TRITYL DERIVATIVES FOR ENHANCING MASS SPECTROMETRY

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

This invention relates to derivatised biopolymers and ions obtainable therefrom. The invention further relates to compounds and solid supports useful for producing the derivatised biopolymers and ions of the invention.

BACKGROUND OF THE INVENTION

Mass spectrometry is a versatile analytical technique possessing excellent detection range and speed of detection with respect to High Performance Liquid Chromatography (HPLC), Gas Chromatography (GC), Infra-Red (IR) and Nuclear Magnetic Resonance (NMR).

However, many biopolymers, such as carbohydrates and proteins, are difficult to analyse using mass spectrometry due to significant difficulties in ionising the biopolymer, even using Matrix Assisted Laser Desorption/Ionisation Time Of Flight (MALDI-TOF) techniques. Despite the considerable resolving power of 2D-PAGE, this technology has fallen far short of the ultimate goal of displaying the whole proteome in a single experiment, as many proteins are resistance to 2D-PAGE analysis (*e.g* those with low or high molecular masses, membrane proteins, proteins with extreme isoelectric points, *etc.*). Many proteins are thus invisible to 2-D PAGE [Cravatt & Sorensen (2000) Current Opinion in Chemical Biology vol. 4, p. 663-668].

There is thus a need for improvements in mass spectrometry analysis of biopolymers.

DISCLOSURE OF THE INVENTION

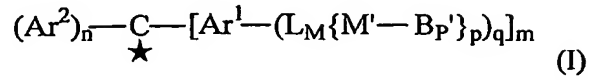
It has now been found that covalent attachment of trityl derivatives to biopolymers can improve the ionisation properties of the biopolymer. The ions (formula (I) below) formed by ionisation of the derivatised biopolymers are particularly suitable for mass spectrometry analysis, and biopolymers derivatised as specified in formulae (IIIa) and (IIIb) below can be readily ionised.

Whereas triphenylmethyl derivatives covalently attached to certain biopolymers (*e.g.* DNA) are known in the prior art [*e.g.* Chem. Soc. Rev. (2003) 32, p. 3-13], the prior art attaches the polymer to the α-triphenylmethyl carbon atom through a non-aromatic linker. In contrast, under the present invention the biopolymer is attached to the α-triarylmethyl carbon atom via an aromatic group adjacent to the central carbon atom. Consequently, ionisation of the prior art derivatives results in separation of the triphenylmethyl derivative and the biopolymer, whereas according to the present invention the biopolymer remains bound to the trityl derivative on ionisation, thereby allowing analysis of the biopolymer by mass spectrometry.

The invention provides methods of forming ions from covalent or ionic compounds and solid substrates.

Derivatised Biopolymers

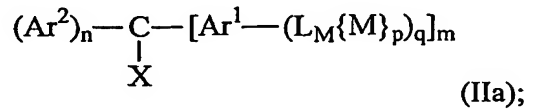
The invention provides a method of forming an ion of formula (I):



(I)

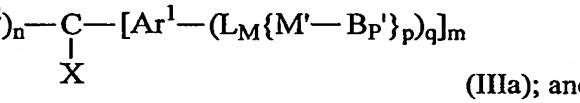
comprising the steps of:

(i) reacting a compound of the formula (IIa):



(IIa);

with a biopolymer, B_P, having at least one group capable of reacting with M to form a covalent linkage, to provide a biopolymer derivative of the formula (IIIa):



(IIIa); and

(ii) cleaving the C—X bond between X and the α-carbon atom of the derivative of formula (IIIa) to form the ion of formula (I);

where:

C★ is a carbon atom bearing a single positive charge or a single negative charge;

X is a group capable of being cleaved from the α-carbon atom to form an ion of formula (I);

M is independently a group capable of reacting with B_P to form the covalent linkage;

B_P' is independently the biopolymer residue of B_P produced on formation of the covalent linkage;

M' is independently the residue of M produced on formation of the covalent linkage;

Ar^1 is independently an aromatic group or an aromatic group substituted with one or more A;

Ar^2 is independently an aromatic group or an aromatic group substituted with one or more A;

optionally wherein (a) two or three of the groups Ar^1 and Ar^2 are linked together by one or more L^5, where L^5 is independently a single bond or a linker atom or group; and/or (b) two or three of the groups Ar^1 and Ar^2 together form an aromatic group or an aromatic group substituted with one or more A;

A is independently a substituent;

L_M is independently a single bond or a linker atom or group;

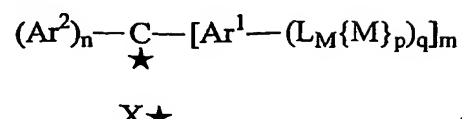
n = 0, 1 or 2 and m = 1, 2, or 3, provided the sum of n+m = 3;

p independently = 1 or more; and

q independently = 1 or more.

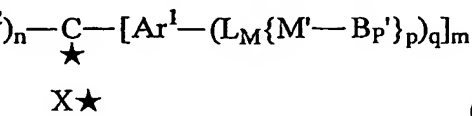
The invention further provides a method of forming an ion of formula (I), comprising the steps of:

(i) reacting a compound of the formula (IIb):



(IIb);

with a biopolymer, B_P, having at least one group capable of reacting with M to form a covalent linkage, to provide a biopolymer derivative of the formula (IIIb):



(IIIb); and

dissociating X★ from the derivative of formula (IIIb), to form the ion of formula (I);

where:

X★ is a counter-ion to C★;

and C★, M, B_P', M', Ar^1, Ar^2, L_M, n, m, p and q are as defined above.

The compounds of formulae (IIa) or (IIb) may optionally be purified after step (i).

The invention also provides biopolymer derivatives of the formula (IIIa) or (IIIb), as defined above. The biopolymer derivatives of the invention have enhanced ionisability with respect to free biopolymer, B_P. Advantageously, the biopolymer derivatives may not require a matrix (*e.g.* as used in MALDI-MS) in order to elicit ionisation, although a matrix may help to enhance ionisation. Preferably, ionisation may be obtained without requiring acid treatment, in particular by direct laser illumination.

The invention also provides ions of formula (I), as defined above. These ions are stabilised by the resonance effect of the aromatic groups Ar^1 and Ar^2. Electron-withdrawing groups, when C★ is an anion, or electron-donating groups, when C★ is a cation, may optionally be provided on Ar^1 and/or Ar^2 to assist this resonance effect. Consequently, the biopolymer derivatives of the invention readily form ions of formula (I) relative to the native biopolymer, B_P.

The ions of formula (I) are generally only ever seen on a mass spectrum with a single charge, which is advantageous since it reduces cluttering of the mass spectrum.

The invention also provides compounds of the formula (IIa) and (IIb), as defined above. As mentioned above, these compounds are useful for forming ions of formula (I). As the difference in the molecular mass of the ions of formula (I) and that of the free biopolymer can be accurately calculated, the derivatised compounds of the invention allow analysis of the biopolymer B_P, which may be otherwise difficult or impossible to analyse using known mass spectrometrical techniques.

Other advantageous features of the compounds of the invention include more uniformity of the signal intensity between different analytes (useful for quantitative studies) and similar desorption properties

between compounds with different, but close, masses, so that techniques such as isotope coded affinity tagging (ICAT) can be employed with the compounds of the invention.

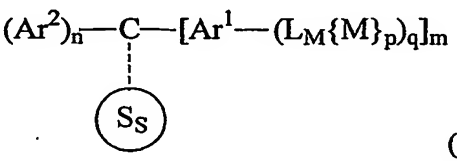
The homogeneous methods of the invention are particularly appropriate for small molecules, *e.g.* amines.

Solid Supports

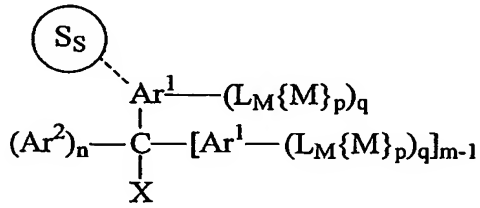
The ions of formula (I) may also be formed using a derivatised solid support.

The invention therefore provides a method of forming an ion of formula (I) comprising the steps of:

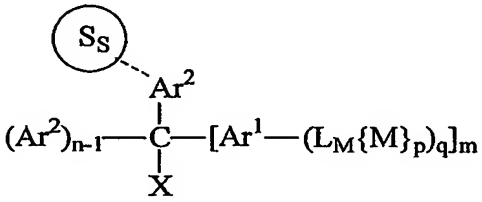
(i) reacting a solid support of formula (IVai), (IVaii), or (IVaiii):



(IVai);

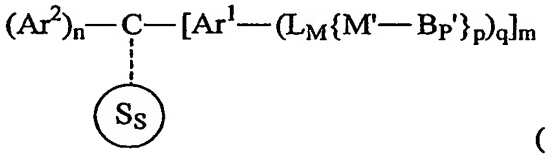


(IVaii);

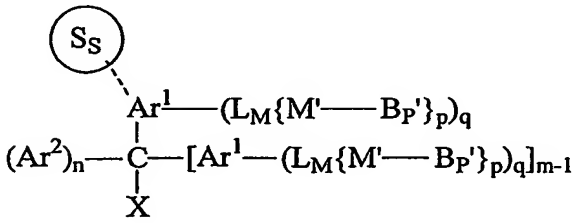


(IVaiii);

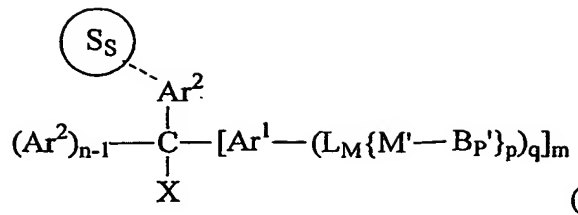
with a biopolymer, B_P, having at least one group capable of reacting with M to form a covalent linkage, to provide a modified solid support of the formula (Vai), (Vaii), or (Vaiii), respectively:



(Vai));



(Vaii);



(Vaiii);

and either:

(iia) for modified solid supports of formula (Vai) cleaving the C-S_S bond between the α-carbon atom of the modified solid support of formula (Vai) and the solid support S_S to form the ion of formula (I);

(iib) for modified solid supports of formula (Vaii), either simultaneously or sequentially, cleaving the C-X bond between X and the α-carbon atom and cleaving the S_S- - -Ar^1 bond between the solid support and the Ar^1 group to form the ion of formula (I); or

(iic) for modified solid supports of formula (Vaiii), either simultaneously or sequentially, cleaving the C-X bond between X and the α-carbon atom and cleaving the S_S- - -Ar^2 bond between the solid support and the Ar^2 group to form the ion of formula (I);

where:

X, Ar^1, Ar^2, B_P', L_M, M, M', n, m, p and q are as defined above;

S_S is a solid support;

C- - -S_S comprises a cleavable bond between C and S_S;

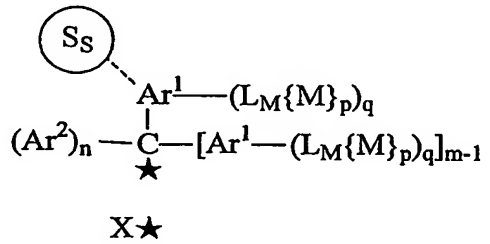
S_S- - -Ar^1 comprises a cleavable bond between Ar^1 and S_S; and

S_S- - -Ar^2 comprises a cleavable bond between Ar^2 and S_S.

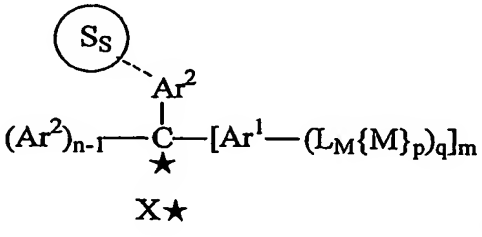
The cleavable bond of C- - -S_S, S_S- - -Ar^1 or S_S- - -Ar^2 may be a covalent, ionic, hydrogen, dipole-dipole or van der Waals bond.

The invention further provides a method of forming an ion of formula (I) comprising the steps of:

(i) reacting a solid support of formula (IVbii) or (IVbiii):

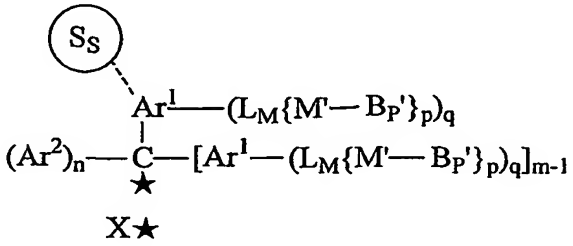


(IVbii);

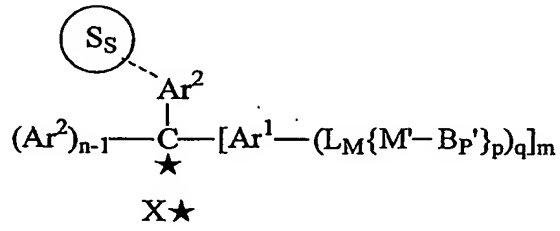


(IVbiii);

with a biopolymer, B_P, having at least one group capable of reacting with M to form a covalent linkage, to provide a modified solid support of the formula (Vbii) or (Vbiii), respectively:



(Vbii);



(Vbiii);

and either:

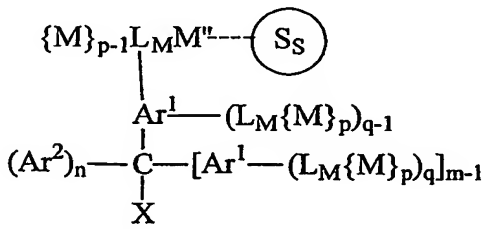
(iia) for modified solid supports of formula (Vbii), either simultaneously or sequentially, dissociating X★ from the derivative of formula (Vbii) and cleaving the S_S- - -Ar^1 bond between the solid support and the Ar^1 group to form an ion of formula (I); or

(iib) for modified solid supports of formula (Vbiii), either simultaneously or sequentially, dissociating X★ from the derivative of formula (Vbiii) and cleaving the S_S- - -Ar^2 bond between the solid support and the Ar^2 group to form an ion of formula (I);

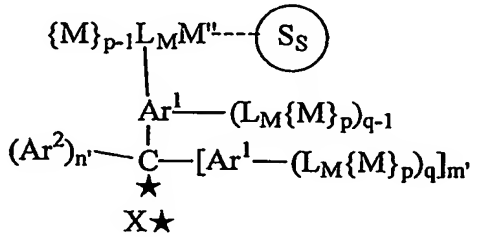
where: X★, Ar^1, Ar^2, B_P', L_M, M, M', n, m, p, q, S_S, C- - -S_S, S_S- - -Ar^1 and S_S- - -Ar^2 are as defined above.

The invention further provides a method of forming an ion of formula (I) comprising the steps of:

(i) reacting a solid support of formula (IVaiv) or (IVbiv):

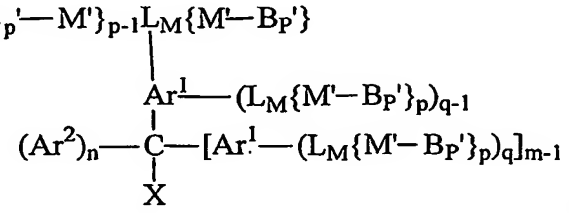


(IVaiv);

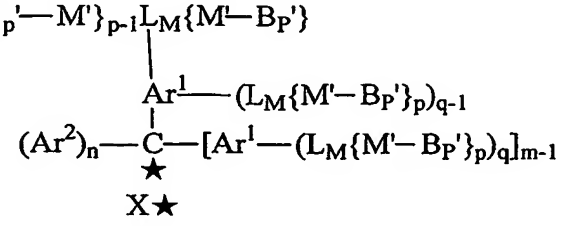


(IVbiv);

with a biopolymer, B_P, having at least one group capable of reacting with M to form a covalent linkage, to provide a modified solid support of the formula (Vaiv) or (Vbiv), respectively:



(Vaiv);



(Vbiv);

and either:

(iia) for modified solid supports of formula (Vaiv), cleaving the C-X bond between X and the α-carbon atom to form the ion of formula (I); or

(iib) for modified solid supports of formula (Vbiv), dissociating X★ from the derivative of formula (Vbiv) to form the ion of formula (I);

where:

X, X★, Ar^1, Ar^2, B_P', L_M, M, M', p, q, n, m, and S_S are as defined above;

M"- - -S_S comprises a bond between M" and S_S; and

M" is the same as M except that S_S is bound to a portion of M which does not form part of M'.

In this embodiment of the invention, the solid support is bound to a part of group M" which does not go on to form the residue M'. Thus, the derivatised biopolymer will be released from the solid support during the derivativisation step and an additional step of cleaving the biopolymer from the solid support is not required.

The modified solid supports of formulae (Vai), (Vaii), (Vaiii), (Vaiv), (Vbii), (Vbiii) or (Vbiv) may optionally be washed after step (i).

The invention also provides solid supports of the formulae (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) and (IVbiv), as defined above. Similarly, the invention provides modified solid supports of the formulae (Vai), (Vaii), (Vaiii), (Vaiv), (Vbii), (Vbiii), and (Vbiv), as defined above.

The heterogeneous methods of the invention are particularly appropriate for synthetic biopolymers, *e.g.* oligonucleotides, peptides and carbohydrates.

Methods of Analysis

The invention also provides a method for analysing a biopolymer, B_P, comprising the steps of:

(i) reacting the biopolymer B_P with a compound of formula (IIa) or (IIb) or a solid support of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv);

(ii) providing an ion of formula (I); and

(iii) analysing the ion of formula (I) by mass spectrometry.

The biopolymer will typically have been obtained using a preparative or analytical process. For example, it may have been purified using various separation methods (*e.g.* 1-dimensional or 2-dimensional, reverse-phase or normal-phase separation, by *e.g.* chromatography or electrophoresis) and the separation may be based on any of a number of characteristics (*e.g.* isoelectric point, molecular weight, charge, hydrophobicity, *etc.*). Typical methods include 2D SDS-PAGE , 2D liquid chromatography (*e.g.* Multidimensional Protein Identification Technology, MudPIT, or 2D HPLC methods). The separation method can preferably interface directly with the mass spectrometer.

Known analytical techniques can thus be adapted or improved by the method of the invention. A particularly preferred method involves 2D-PAGE of a biopolymer, or mixture of biopolymers, selection of a spot of interest in the electrophoretogram, and then derivatisation and analysis of that spot using the techniques of the invention. The biopolymer may be proteolytically digested prior to its analysis (typically within the PAGE gel, but optionally digested after extraction from the gel) and/or may itself be the product of a proteolytic digest.

The invention also provides, in a method for analysing a biopolymer, B_P, the improvement consisting of: (i) reacting a biopolymer, B_P with a compound of formula (IIa) or (IIb) or a solid support of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv); (ii) providing an ion of formula (I); and (iii) analysing the ion by mass spectrometry.

Typically, the analysis by mass spectrometry is carried out in a spectrometer which is suitable for MALDI-TOF spectrometry.

In the spectrometer, the ion source may be a matrix-assisted laser desorption ionisation (MALDI), an electrospray ionisation (ESI) ion source, a Fast-Atom Bombardment (FAB) ion source. Preferably, the ion source is a MALDI ion source. The MALDI ion source may be traditional MALDI source (under vacuum) or may be an atmospheric pressure MALDI (AP-MALDI) source. MALDI is a preferred ionisation method, although the use of a matrix is generally not required

In the spectrometer, the mass analyser may be a time of flight (TOF), quadrupole time of flight (Q-TOF), ion trap (IT), quadrupole ion trap (Q-IT), triple quadrupole (QQQ) Ion Trap or Time-Of-Flight Time-Of-Flight (TOFTOF) or Fourier transform ion cyclotron resonance (FTICR) mass analyser. Preferably, the mass analyser is a TOF mass analyser.

Preferably, the mass spectrometer is a MALDI-TOF mass spectrometer.

Further Embodiments

M' bound to B_P' by a non-covalent linker

The above-mentioned embodiments of the invention may also be provided in which M' is bound to B_P' by a non-covalent bond. All the other features of the invention are the same except the groups which relate to the non-covalent bond between M' and B_P'.

The non-covalent bond may be direct between M' and B_P' or may be provided by one or more binding groups present on M' and/or B_P'.

Preferred non-covalent bonds are those having an association constant (K_a) of at least 10^{14} M^{-1}, preferably about 10^{15} M^{-1}.

In preferred embodiment, one of M' and B_P' will have a binding group comprising biotin, and the other of M' and B_P' will have a binding group comprising avidin or streptavidin.

Preferably, when the compounds of the invention comprise a non-covalent bond between M' and B_P' and a cleavable bond between C and S_S, Ar^1 and S_S, or Ar^2 and S_S, these bonds are differentially cleavable. More preferably, the non-covalent bond between M' and B_P' is not cleaved under conditions which the cleavable bond between C and S_S, Ar^1 and S_S, or Ar^2 and S_S, as appropriate, is cleaved.

L_M bound to Ar^1 by more than one bond

The above-mentioned embodiments of the invention may also be provided in which L_M is bound to Ar^1 by more than one covalent bond (*e.g.* 2 or 3 bonds) which are either single, double or triple covalent bonds, or one or more multiple bonds (*e.g.* double or triple covalent bonds). All the other features of the invention are the same except the groups which relate to the bond or bonds between Ar^1 and L_M.

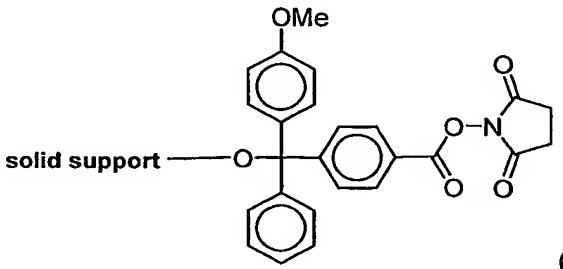
Ionisation of Compounds other than Biopolymers

In addition to biopolymers, the present invention may be used for ionising any molecule or complex of molecules which requires mass spectrum analysis. Thus, the above-mentioned embodiments of the invention may also be provided in which B_P is replaced by any molecule or complex having at least one group capable of reacting with M to form a covalent linkage. All the other features of the invention are the same, except group M is group capable of reacting with the molecule to be analysed.

Examples of other molecules which may be analysed in the present invention include non-biological polymers (*e.g.* synthetic polyesters, polyamides and polycarbonates), petrochemicals and small molecules (*e.g.* alkanes, alkenes, amines, alcohols, esters and amides). Amines are particularly preferred.

Examples of complexes which may be analysed in the present invention include double- and triple-stranded RNA, DNA and/or peptide nucleic acid (PNA) complexes, enzyme/substrate complexes, multimeric proteins (*e.g.* dimers, trimers, tetramers, pentamers, *etc.*), virions, *etc.*

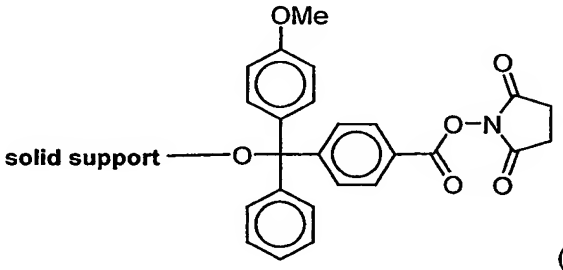
Preferably, when the compound to be ionised is not a biopolymer, all embodiments of the invention (including products of formulae (I), (IIa), (IIb), (IIIa), (IIIb), (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii), (IVbiv), (Vai), (Vaii), (Vaiii), (Vaiv), (Vbii), (Vbiii) and (Vbiv), methods of forming an ion of formula (I) and methods of analysis) involving or relating to the compound of formula (XI) are disclaimed.



(XI).

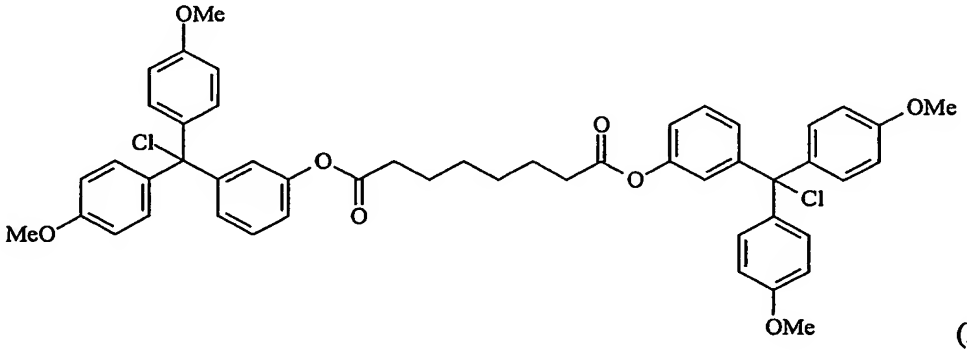
Disclaimers

Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XI) are disclaimed



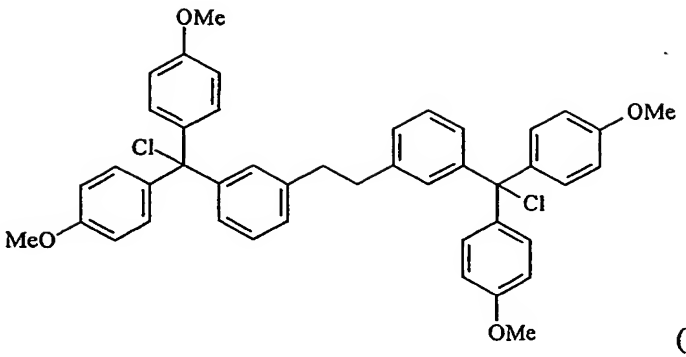
(XI).

Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XIa) are disclaimed



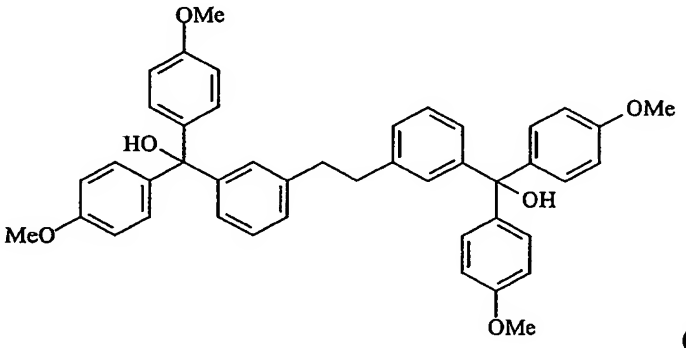
(XIa).

Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XIb) are disclaimed.



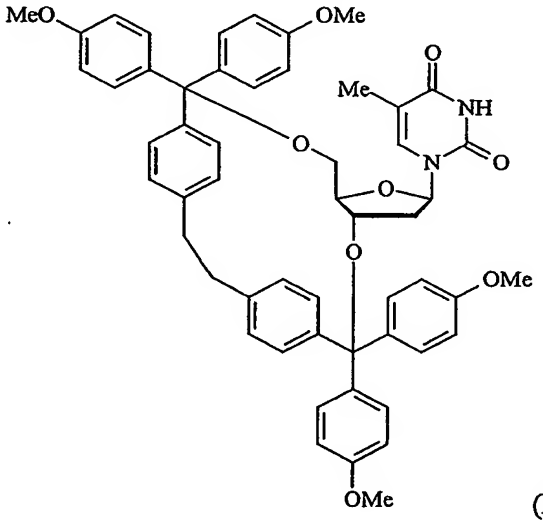
(XIb).

Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XIc) are disclaimed



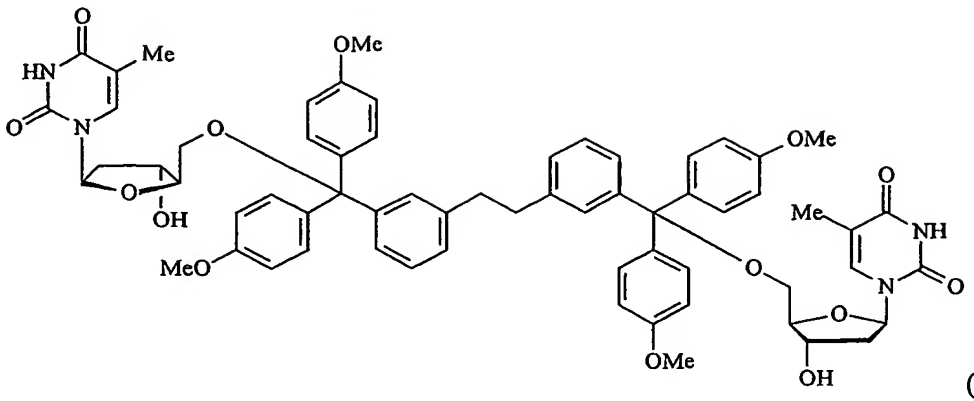
(XIc).

Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XId) are disclaimed



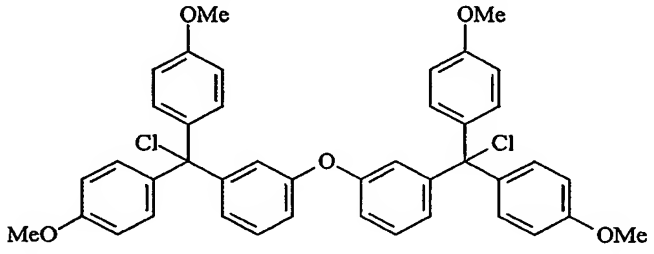
(XId).

Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XIe) are disclaimed



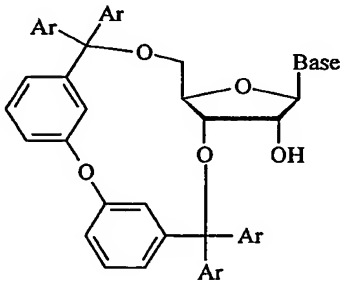
(XIe).

Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XIe) are disclaimed



(XIf).

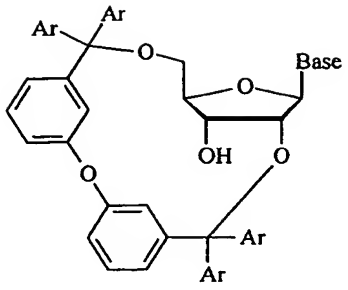
Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XIg-j) are disclaimed



Ar = p-anisyl

Formula	Base
XIg	Uridine
XIh	N^4-benzoyl-cytidine
XIi	N^6-benzoyl-adenosine
XIj	N^2-phenylacetyl-guanosine

Preferably, all embodiments of the invention (including products of formulae (I) and (IIa)) involving or relating to the compound of formula (XIk-n) are disclaimed



Ar = p-anisyl

Formula	Base
XIk	Uridine
XIl	N^4-benzoyl-cytidine
XIm	N^6-benzoyl-adenosine
XIn	N^2-phenylacetyl-guanosine

Preferred Embodiments

Definition of C★

Preferably, C★ bears a single positive charge such that ions of the invention are cations and the ion of formula (I) has the following structure:

$$(Ar^2)_n—\underset{\oplus}{C}—[Ar^1—(L_M\{M'—B_P'\}_p)_q]_m$$

and the compounds of formulae (IIb), (IIIb), (IVbii), (IVbiii), (IVbiv), (Vbii), (Vbiii) and (Vbiv) have the structures disclosed in table 1.

n, m, p and q

For the purposes of compounds of the invention having n-1 groups Ar^2, n may not be less than 1.

Preferably n = 2 and m = 1.

Preferably p = 1, 2 or 3. Preferably p = 1.

Preferably q = 1, 2 or 3. Preferably q = 1.

Preferably n = 2, m = 1, p = 1 and q = 1. The ion of formula (I) thus has the structure:

$$Ar^2—\underset{\star}{\overset{Ar^2}{\overset{|}{C}}}—Ar^1—L_MM'—B_P'$$ or more preferably $$Ar^2—\underset{\oplus}{\overset{Ar^2}{\overset{|}{C}}}—Ar^1—L_MM'—B_P'$$

and the compounds of formulae (IIa), (IIb), (IIIa), (IIIb), (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii), (IVbiv), (Vai), (Vaii), (Vaiii), (Vaiv), (Vbii), (Vbiii) and (Vbiv) have the structures disclosed in table 2.

Biopolymers

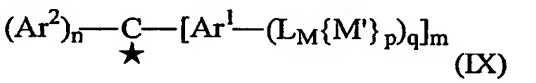
The term 'biopolymer' includes polymers found in biological samples, including polypeptides, polysaccharides, and polynucleotides (*e.g.* DNA or RNA). Polypeptides may be simple copolymers of amino acids, or they may include post-translational modifications *e.g.* glycosylation, lipidation, phosphorylation, *etc.* Polynucleotides may be single-stranded (in whole or in part), double-stranded (in whole or in part), DNA/RNA hybrids, *etc.* RNA may be mRNA, rRNA or tRNA.

Advantageous biopolymers are those which do not readily form a molecular ion in known MALDI-TOF MS techniques, especially those which do not form a molecular ion on illumination of laser light at 340 nm.

Biopolymers for use in the invention comprise two or more monomers, which may be the same or different as each other. Preferred biopolymers comprise at least *pp* monomers, where *pp* is 5 or more (*e.g.* 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250). More preferred biopolymers comprise *ppp* or fewer monomers where *ppp* is 300 or less (*e.g.* 200, 100, 50).

Biopolymers may have a molecular mass of at least *qq* kDa, where *qq* = 0.5 or more (*e.g.* 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, *etc.*). Preferred biopolymers are those having a molecular mass within the range of detection of a mass spectrometer. More preferred biopolymers have a molecular mass of *qqq* kDa or less, where *qqq* is 30 or less (*e.g.* 20, 10, 5).

Preferably, the mass, m(IX), of the fragment (IX)



(IX)

of the cation of formula (I) is significantly less than the mass, m(B_P'), of the biopolymer residue B_P'. For example the ratio m(B_P') / m(IX) is preferably more than *nn*, where *nn* is at least 2 (*e.g.* 3, 4, 5, 10, 100, 1000, *etc.*).

The invention is suitable for use with purified biopolymers or mixtures of biopolymers. For example, a pure recombinant protein could be derivatised and analysed by MS, or biopolymers within a cellular lysate or extract could be derivatives and then analysed.

Preferred biopolymers are polypeptides. Particularly preferred biopolymers are polypeptides formed after proteolytic digestion of a protein.

Biopolymers bound to solid supports

In preferred embodiments of the invention the biopolymer is bound to a solid support such that it is cleavable from the solid support at least once it has been derivatised by a compound of the invention. B_P is thus derivatised *in situ* while bound to the support, and is then released. As the biopolymer is bound to the solid support, this aspect of the invention is particular relevant to methods involving compounds of formulae (IIa) and (IIb).

The biopolymer may be bound to the solid support by a covalent, ionic, hydrogen, dipole-dipole or van der Waals bond (also known as a dispersion bond or a London forces bond). The covalent, ionic, hydrogen, dipole-dipole or van der Waals bond may be direct between the biopolymer and the solid support or may be provided by one or more binding groups present on the biopolymer and/or solid support. Preferred groups are non-covalent groups.

Examples of groups which can form these types of bond, and methods for cleaving these types of bond, are set out below in connection with C---S_S bonds, *etc.*

In a particularly preferred embodiment, the solid support is provided with $-(NMe_3)^+$ binding groups and the biopolymer has a net negative charge, or *vice versa* (*i.e.* the $-(NMe_3)^+$ is on the biopolymer). In other preferred embodiments, the solid support is provided with anions such as carboxylate, phosphate or sulphate, or anions formed from acid groups, and the biopolymer (*e.g.* a histone) has a net positive charge, or *vice versa*.

Reactivity with group M

The biopolymers have at least one reactive group capable of reacting with M to form a covalent linkage. Such groups typically include naturally occurring groups and groups formed synthetically on the biopolymer.

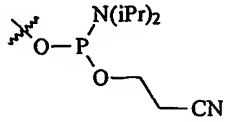
Naturally occurring groups include lipid groups of lipoproteins (*e.g.* myristoyl, glycosylphosphatidylinositol, ethanolamine phosphoglycerol, palmitate, stearate, S- or N- or O-acyl groups, lipoic acid, isoprenyl, geranylgeranyl, farnesyl, *etc.*), amide, carbohydrate groups of *N*- and *O*- glycoproteins, amine groups (*e.g.* on lysine residues or at the N-terminus of a protein), hydroxyl (*e.g.* in β-hydroxyaspartate, β-hydroxyasparagine, 5-hydroxylysine, ¾-hydroxyproline), thiol, sulfhydryl, phosphoryl, sulfate, methyl, acetyl, formyl (*e.g.* on N-terminal methionines from prokaryotes), phenyl, indolyl, guanidyl, hydroxyl, phosphate, methylthio, ADP-ribosyl *etc.*

The reactive group is bound to the biopolymer by one or more covalent bonds (*e.g.* 2 or 3 bonds), which are either single, double or triple covalent bonds (preferably single bonds). Preferably, the reactive group is bound to the biopolymer by one single bond.

Groups which may be formed naturally or synthetically on the biopolymer and which are bound to the biopolymer by one bond include: $-NR_2$ *e.g.* -NHR, especially $-NH_2$; -SR *e.g.* -SH; -OR *e.g.* -OH; -B(R)Y; $-BY_2$; $-C(R)_2Y$; $-C(R)Y_2$; $-CY_3$; -C(=Z)Y *e.g.* -C(=O)Y; -Z-C(=Z)Y; -C(=Z)R *e.g.* -C(=Z)H, especially -C(=O)H; -C(R)(OH)OR; $-C(R)(OR)_2$; -S(=O)Y; -Z-S(=O)Y; $-S(=O)_2Y$; $-Z-S(=O)_2Y$; $-S(=O)_3Y$; $-Z-S(=O)_3Y$; -P(=Z)(ZR)Y *e.g.* -P(=O)(OH)Y; $-P(=Z)Y_2$; -Z-P(=Z)(ZR)Y; $-Z-P(=Z)Y_2$; -P(=Z)(R)Y *e.g.* -P(=O)(H)Y; -Z-P(=Z)(R)Y; or -N=C(=Z) *e.g.* -N=C(=O).

Another group which may be formed naturally or synthetically on the biopolymer and which is bound to the biopolymer by one bond is –CN.

Other groups which may be formed naturally or synthetically on the biopolymer and which are bound to the biopolymer by one bond are: -P(ZR)Y *e.g.* -P(OH)Y; -PY_2; -Z-P(ZR)Y; -Z-PY_2; -P(R)Y *e.g.* -P(H)Y; -Z-P(R)Y. A particularly preferred group is -Z-P(ZR)Y, especially a phosphoramidite group:



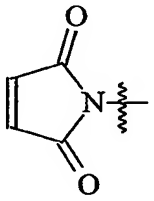
Another example of a group which may be formed naturally or synthetically on the biopolymer and which is bound to the biopolymer by one bond is –Y. In particular, when the reactive group is halo (especially iodo), the reactive group may be bound to an aliphatic or aromatic carbon.

Groups which may be formed synthetically on the biopolymer and which are bound to the biopolymer by two bonds include -N(R)- *e.g.* -NH-; -S-; -O-; -B(Y)-; -C(R)(Y)-; -CY_2-; -C(=O)-; -C(OH)(OR)-; -$C(OR)_2$-.

Groups which may be formed synthetically on the biopolymer and which are bound to the biopolymer by three bonds include .

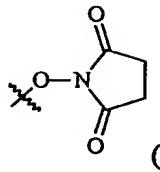
Preferred groups include nucleophilic groups, either natural or synthetic, *e.g.*: -NR_2 *e.g.* -NHR, especially -NH_2; -SR *e.g.* -SH; -OR *e.g.* -OH; -N(R)- *e.g.* -NH-; -S-; and -O-. The groups -NH_2, -SH and -OH are particularly preferred.

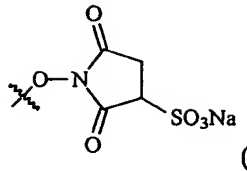
Another preferred reactive group is maleimidyl:

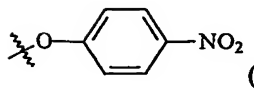


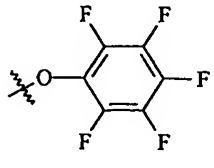
Y is independently a leaving group, including groups capable of leaving in an SN_2 substitution reaction or being eliminated in an addition-elimination reaction with the reactive group of the biopolymer B_P.

Preferred examples of Y include halogen (preferably iodo), C_{1-8}hydrocarbyloxy (*e.g.* C_{1-8}alkoxy), C_{1-8}hydrocarbyloxy substituted with one or more A, C_{1-8}heterohydrocarbyloxy, C_{1-8}heterohydrocarbyloxy substituted with one or more A, mesyl, tosyl, pentafluorophenyl, -O-succinimidyl (formula VII) or a sulfo sodium salt thereof (sulfoNHS – formula VIIa), -S-succinimidyl, or phenyloxy substituted with one or more A *e.g.* p-nitrophenyloxy (formula VIII) or pentafluorophenoxy (formula VIIIa).

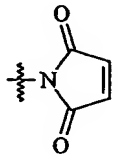

(VII)

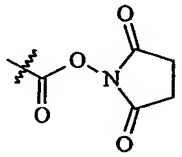

(VIIa)

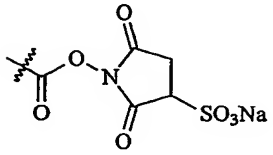

(VIII).

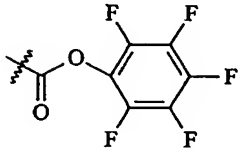

(VIIIa)

Thus, preferred reactive group on the biopolymer are:









Other preferred examples of Y include –ZR. Particularly preferred examples of Y are –ZH (*e.g.* –OH or $–NH_2$) and -Z-C_{1-8}alkyl groups such as -NH-C_{1-8}alkyl groups (*e.g.* -NHMe) and -O-C_{1-8}alkyl groups (*e.g.* -O-t-butyl). Thus, preferred reactive groups are -C(O)-NH-C_{1-8}alkyl and -C(O)-O-C_{1-8}alkyl (*e.g.* -C(O)-O-t-butyl).

Other preferred examples of Y include –Z-ZR. Particularly preferred examples include –NR-NR_2, especially –NH-NH_2, and –ONR_2, especially –O-NH_2.

Z is independently O, S or N(R). Preferred (=Z) is (=O).

R is independently H, C_{1-8}hydrocarbyl (*e.g.* C_{1-8}alkyl) or C_{1-8}hydrocarbyl substituted with one or more A.

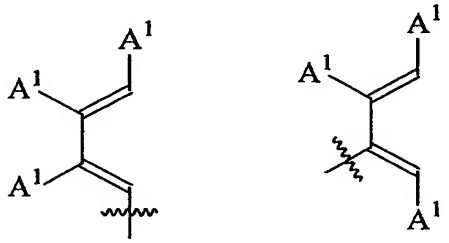
R is preferably H.

Other preferred reactive groups include -C(=O)Y, especially -C(=O)-O-succinimidyl and -C(=O)-O-(p-nitrophenyl).

In a further embodiment, the reactive group may be $-Si(R)_2-Y$, with Y being halo (*e.g.* chloro) being especially preferred. Preferred groups R in this embodiment are C_{1-8}alkyl, especially methyl. A particularly preferred reactive group in this embodiment is $-Si(Me)_2Cl$.

Other groups which may be formed naturally or synthetically on the biopolymer include groups capable of reacting in a cycloaddition reaction, especially a Diels-Alder reaction.

In the case of Diels-Alder reactions, the reactive group on the biopolymer is either a diene or a dienophile. Preferred diene groups are

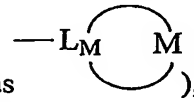


and multivalent derivatives formally formed by removal of one or more hydrogen atoms, where A^1 is $-R^1$ or $-Z^1R^1$, where R^1 and Z^1 are defined below.

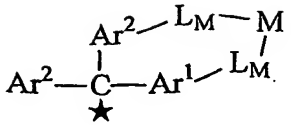
Preferred dienophile groups are $-CR^1=CR^1_2$, $-CR^1=C(R^1)A^2$, $-CA^2=CR^1_2$, $-CA^2=C(R^1)A^2$ or $-CA^2=CA^2_2$, and multivalent derivatives formally formed by removal of one or more hydrogen atoms, where R^1 is defined below and A^2 is independently halogen, trihalomethyl, $-NO_2$, -CN, $-N^+(R^1)_2O^-$, $-CO_2H$, $-CO_2R^1$, $-SO_3H$, $-SOR^1$, $-SO_2R^1$, $-SO_3R^1$, $-OC(=O)OR^1$, -C(=O)H, $-C(=O)R^1$, $-OC(=O)R^1$, , $-OC(=O)NR^1_2$, $-N(R^1)C(=O)R^1$, $-C(=S)NR^1_2$, $-NR^1C(=S)R^1$, $-SO_2NR^1_2$, $-NR^1SO_2R^1$, $-N(R^1)C(=S)NR^1_2$, or $-N(R^1)SO_2NR^1_2$, where R^1 is defined below. A particularly preferred dienophile group is maleimidyl.

Group M

The group M is capable of reacting with the reactive group of the biopolymer, B_P, to form a covalent linkage. [Group 'M' is shown as 'AFG' in the drawings].

The group M is bound to L_M by one or more covalent bonds (*e.g.* 2 or 3 bonds, especially 2 such as ), which are either single, double or triple covalent bonds (preferably single bonds). Preferably, M is bound to L_M by one single bond.

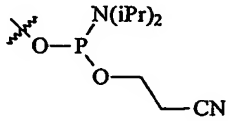
Alternatively, or in addition, M is bound by more than one L_M, such L_M either being attached to the same or different Ar^1 or Ar^2. In a preferred embodiment M is bound by more than one L_M from different Ar^1 or Ar^2, *e.g.*:



Examples of group M bound to L_M by one bond include $-NR_2$ *e.g.* -NHR, especially $-NH_2$; -SR *e.g.* -SH; -OR *e.g.* -OH; -B(R)Y; $-BY_2$; $-C(R)_2Y$; $-C(R)Y_2$; $-CY_3$; -C(=Z)Y *e.g.* -C(=O)Y; -Z-C(=Z)Y; -C(=Z)R *e.g.* -C(=Z)H, especially -C(=O)H; -C(R)(OH)OR; $-C(R)(OR)_2$; -S(=O)Y; -Z-S(=O)Y; $-S(=O)_2Y$; $-Z-S(=O)_2Y$; $-S(=O)_3Y$; $-Z-S(=O)_3Y$; -P(=Z)(ZR)Y *e.g.* -P(=O)(OH)Y; $-P(=Z)Y_2$; -Z-P(=Z)(ZR)Y; $-Z-P(=Z)Y_2$; -P(=Z)(R)Y *e.g.* -P(=O)(H)Y; -Z-P(=Z)(R)Y; or -N=C(=Z) *e.g.* -N=C(=O).

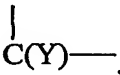
Another example of a group M bound to L_M by one bond is –CN.

Other examples of group M bound to L_M by one bond are -P(ZR)Y *e.g.* -P(OH)Y; $-PY_2$; -Z-P(ZR)Y; $-Z-PY_2$; -P(R)Y *e.g.* -P(H)Y; -Z-P(R)Y. A particularly preferred group M is -Z-P(ZR)Y, especially a phosphoramidite group:



Another example of group M bound to L_M by one bond is –Y. In particular, when group M is halo (especially iodo), M may be bound to an aliphatic or aromatic carbon. When M is halo (*e.g.* iodo) and is bound to an aromatic carbon, L_M may, for example, be a single bond.

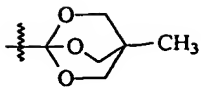
Examples of group M bound to L_M by two bonds include -N(R)- *e.g.* -NH-; -S-; -O-; -B(Y)-; -C(R)(Y)-; $-CY_2-$; -C(=O)-; -C(OH)(OR)-; $-C(OR)_2-$.

Examples of group M bound to L_M by three bonds include  .

Preferred groups M include electrophilic groups, especially those susceptible to SN_2 substitution reactions, addition-elimination reactions and addition reactions, *e.g.* -B(R)Y; $-BY_2$; $-C(R)_2Y$; $-C(R)Y_2$; $-CY_3$; -C(=Z)Y *e.g.* -C(=O)Y; -Z-C(=Z)Y; -C(=Z)R *e.g.* -C(=Z)H, especially -C(=O)H; -C(R)(OH)OR; $-C(R)(OR)_2$; -S(=O)Y; -Z-S(=O)Y; $-S(=O)_2Y$; $-Z-S(=O)_2Y$; $-S(=O)_3Y$; $-Z-S(=O)_3Y$; -P(=Z)(ZR)Y *e.g.* -P(=O)(OH)Y; $-P(=Z)Y_2$; -Z-P(=Z)(ZR)Y; $-Z-P(=Z)Y_2$; -P(=Z)(R)Y *e.g.* -P(=O)(R)Y; -Z-P(=Z)(H)Y; -N=C(=Z) *e.g.* –N=C(=O); -B(Y)-; -C(R)(Y)-; $-CY_2-$; -C(=O)-; $-C(OH)(OR)-$; $-C(OR)_2-$; or .

Another preferred electrophilic group M is –CN.

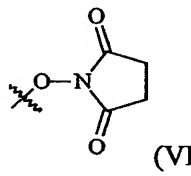
Still further preferred examples of group M are orthoesters, *e.g.* $-C(OR)_3$. In a preferred embodiment, the R groups are linked together to form a hydrocarbyl group, *e.g.* a C_{1-8}alkyl group. A preferred example of group M in this embodiment is:



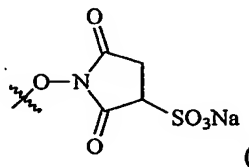
Another preferred group M is maleimido.

Y, Z and R are defined as above. Preferred Y groups when present on M are those capable of leaving in an SN_2 substitution reaction or being eliminated in an addition-elimination reaction with the reactive group of the biopolymer B_P.

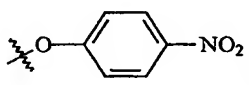
Preferred examples of Y include halogen (preferably iodo), C_{1-8}hydrocarbyloxy (*e.g.* C_{1-8}alkoxy), C_{1-8}hydrocarbyloxy substituted with one or more A, C_{1-8}heterohydrocarbyloxy, C_{1-8}heterohydrocarbyloxy substituted with one or more A, mesyl, tosyl, pentafluorophenyl, -O-succinimidyl (formula VII) or a sulfo sodium salt thereof (sulfoNHS – formula VIIa), -S-succinimidyl, or phenyloxy substituted with one or more A *e.g.* p-nitrophenyloxy (formula VIII) or pentafluorophenoxy (formula VIIIa).



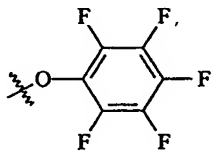
(VII)



(VIIa)

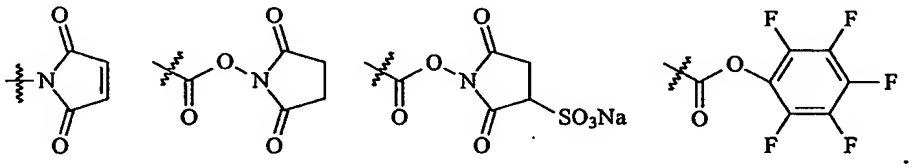


(VIII).



(VIIIa)

Thus, preferred groups M are:



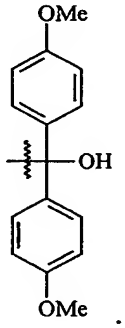
Other preferred examples of Y include –ZR. Particularly preferred examples of Y are –ZH (*e.g.* –OH or $–NH_2$) and -Z-C_{1-8}alkyl groups such as -NH-C_{1-8}alkyl groups (*e.g.* -NHMe) and -O-C_{1-8}alkyl groups (*e.g.* -O-t-butyl). Thus, preferred groups M are -C(O)-NH-C_{1-8}alkyl (*e.g.* -C(O)NHMe) and -C(O)-O-C_{1-8}alkyl (*e.g.* -C(O)-O-t-butyl).

Other preferred examples of Y include –Z-ZR. Particularly preferred examples include $–NR-NR_2$, especially $–NH-NH_2$, and $–ONR_2$, especially $–O-NH_2$.

Particularly preferred groups M include -C(=O)Y, especially -C(=O)-O-succinimidyl and -C(=O)-O-(p-nitrophenyl).

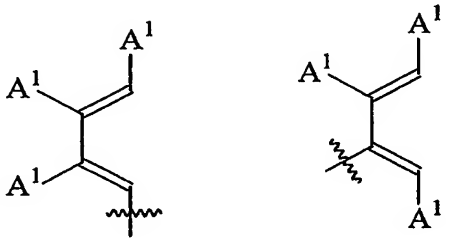
In a further embodiment, M may be $-Si(R)_2-Y$, with Y being halo (*e.g.* chloro) being especially preferred. Preferred groups R in this embodiment are C_{1-8}alkyl, especially methyl. A particularly preferred group M in this embodiment is $-Si(Me)_2Cl$.

In a further embodiment, M may be $-C(Ar^2)_2X$. Preferred groups Ar^2 and X are set out below. In this embodiment it is preferred that L_M is a bond. A particularly preferred group M in this embodiment is:



Other groups M include groups capable of reacting in a cycloaddition reaction, especially a Diels-Alder reaction.

In the case of Diels-Alder reactions, the reactive group on the biopolymer is either a diene or a dienophile. Preferred diene groups are



and multivalent derivatives formally formed by removal of one or more hydrogen atoms, where A^1 is $-R^1$ or $-Z^1R^1$, where R^1 and Z^1 are defined below.

Preferred dienophile groups are $-CR^1=CR^1_2$, $-CR^1=C(R^1)A^2$, $-CA^2=CR^1_2$, $-CA^2=C(R^1)A^2$ or $-CA^2=CA^2_2$, and multivalent derivatives formally formed by removal of one or more hydrogen atoms, where R^1 is defined below and A^2 is independently halogen, trihalomethyl, $-NO_2$, -CN, $-N^+(R^1)_2O^-$, $-CO_2H$, $-CO_2R^1$, $-SO_3H$, $-SOR^1$, $-SO_2R^1$, $-SO_3R^1$, $-OC(=O)OR^1$, $-C(=O)H$, $-C(=O)R^1$, $-OC(=O)R^1$, , $-OC(=O)NR^1_2$, $-N(R^1)C(=O)R^1$, $-C(=S)NR^1_2$, $-NR^1C(=S)R^1$, $-SO_2NR^1_2$, $-NR^1SO_2R^1$, $-N(R^1)C(=S)NR^1_2$, or $-N(R^1)SO_2NR^1_2$, where R^1 is defined below. A particularly preferred dienophile group is maleimidyl.

Preferred examples of group M are shown in figures 11A and 11B.

Matching B_P and M

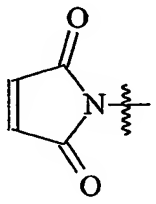
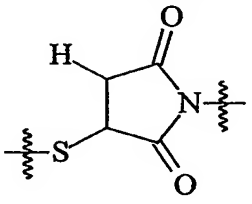
The reactive group on the biopolymer [shown as 'F' in the drawings] and the group M [shown as 'AFG' in the drawings] must be dependently selected in order to form the covalent linkage. For

example, where the biopolymer includes the groups $-NH_2$, -OH or -SH, M will typically be -B(R)Y; $-BY_2$; $-C(R)_2Y$; $-C(R)Y_2$; $-CY_3$; -C(=Z)Y *e.g.* -C(=O)Y; -Z-C(=Z)Y; -C(=Z)R *e.g.* -C(=Z)H, especially -C(=O)H; -C(R)(OH)OR; $-C(R)(OR)_2$; -S(=O)Y; -Z-S(=O)Y; $-S(=O)_2Y$; $-Z-S(=O)_2Y$; $-S(=O)_3Y$; $-Z-S(=O)_3Y$; -P(=Z)(ZR)Y *e.g.* -P(=O)(OH)Y; $-P(=Z)Y_2$; -Z-P(=Z)(ZR)Y; $-Z-P(=Z)Y_2$; -P(=Z)(R)Y *e.g.* -P(=O)(H)Y; -Z-P(=Z)(R)Y; -N=C(=Z) *e.g.* -N=C(=O); -B(Y)-; -C(R)(Y)-; $-CY_2-$; -C(=O)-; -C(OH)(OR)-; $-C(OR)_2-$; or $-\overset{|}{C}(Y)-$. M may also be –CN.

In a preferred embodiment, one of the reactive group on the biopolymer and group M is a maleimidyl and the other will be a –SH group.

Alternatively, when the covalent linkage is to be formed by a Diels Alder reaction, one of the reactive group on the biopolymer and group M will typically be a diene and the other will be a dienophile.

Preferred covalent linkages are those produced through the reaction of the following groups:

M	Group on B_P	Obtained Linkage M'-B_P'
–C(=O)-O-succinimidyl [*i.e.* carboxy-NHS]	$-NH_2$	-CO-NH-
-C(=O)-O-(p-nitrophenyl)	$-NH_2$	-CO-NH-
-C(=O)-pentafluorophenyl	$-NH_2$	-CO-NH-
Biotin	avidin / streptavidin	biotin-(strept)avidin
	-SH	
-N=C=S (isothiocyanate)	$-NH_2$	-NH-CS-NH-

The covalent residue M'-B_P' is the reaction product of M and B_P. B_P' will generally be the same as B_P except that instead of the reactive group, B_P' will have a residue of the reactive group covalently bound to the residue M'. Depending on the choice of the reactive group and the choice of M, M' and the residue of the reactive group will typically form linkages, in the orientation L_M-M'-B_P', including $-C(R)_2Z-$, $-ZC(R)_2-$, -C(=Z)Z-, -ZC(=Z)-, -ZC(=Z)Z-, -C(OH)(R)Z-, -ZC(OH)(R)-, -C(R)(OR)Z-, -ZC(R)(OR)-, -C(R)(OR)Z-, -ZC(R)(OR)-, -S(=O)Z-, -ZS(=O)-, -ZS(=O)Z-, $-S(=O)_2Z-$, $-ZS(=O)_2-$, $-ZS(=O)_2Z-$, $-S(=O)_3Z-$, $-ZS(=O)_3-$, $-ZS(=O)_3Z-$, -P(=Z)(ZR)Z-, -ZP(=Z)(ZR)-, -ZP(=Z)(ZR)Z-, -P(=Z)(R)Z-, -ZP(=Z)(R)-, -ZP(=Z)(R)Z-, -NH-C(=Z)-Z-, where Z and R are as defined above.

Group M''

M'' is the same as M except that the group S_S is bound to a portion of M which does not form part of M'. Thus, M'' is a residue of M formable by the conjugation of M and S_S. However, M'' need not necessarily be formed by the conjugation of M and S_S.

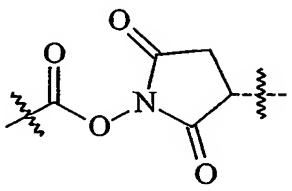
M''- - -S_S comprises a covalent, ionic, dipole-dipole, hydrogen, or van der Waals bond. The covalent, ionic, hydrogen, dipole-dipole or van der Waals bond may be direct between M'' and S_S or may be provided by one or more binding groups present on M'' and/or S_S.

Examples of groups which can form these types of bond, and methods for cleaving these types of bond, are set out below in connection with C- - -S_S bonds, *etc.*

This embodiment of the invention is advantageous, since the derivatisation of the biopolymer will also release the derivatised biopolymer from the solid support. Thus, an additional step of cleaving the biopolymer from the solid support is not required.

Preferred groups M'' are groups M having a leaving group, wherein the group S_S is bound to the leaving group, *e.g.* groups M mentioned above having a leaving group Y, wherein the group S_S is bound to the leaving group Y.

A particularly preferred group M'' is:



L_M

Where the group L_M is a linker atom or group, it has a sufficient number of linking covalent bonds to link L_M to the group Ar^1 by a single covalent bond (or more, as appropriate) and to link L_M to the p instances of M (or M', as appropriate) groups (which may be attached to L_M by one or more bonds).

The group L_M may be directly bound to the aromatic part of Ar^1, bound to one or more of the substituents A of Ar^1, or both. Preferably, L_M is bound directly to the aromatic part of Ar^1.

In an alternative embodiment, L_M may be bound to L_5.

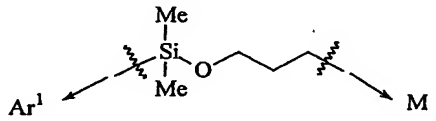
When L_M is a linker atom, preferred linker atoms are O or S, particularly O.

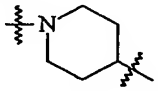
When L_M is a linker group, preferred linker groups, in the orientation Ar^1-$(L_M\{M\}_p)_q$ or Ar^1-$(L_M\{M'\}_p)_q$, as appropriate, are -E^M-, -$(D^M)_t$-, -$(E^M$-$D^M)_t$-, -$(D^M$-$E^M)_t$-, -E^M-$(D^M$-$E^M)_t$- or -D^M-$(E^M$-$D^M)_t$-, where a sufficient number of linking covalent bonds, in addition to the covalent bonds at the chain termini shown, are provided on groups E^M and D^M for linking the p instances of M (or M') groups.

D^M is independently C_{1-8}hydrocarbylene or C_{1-8}hydrocarbylene substituted with one or more A. Preferred D^M are C_{1-8}alkylene, C_{1-8}alkenylene and C_{1-8}alkynylene, especially C_{1-8}alkylene and C_{1-8}alkynylene, each optionally substituted with one or more A (preferably unsubstituted). A preferred substituent A is ^2H. Preferred L_M in the orientation Ar^1-$(L_M\{M\}_p)_q$ or Ar^1-$(L_M\{M'\}_p)_q$, as appropriate, are: $-CH_2CH_2-$; $-C\equiv C-CH_2CH_2CH_2-$; $-(CH_2)_5-$; $-CD_2CD_2CH_2CH_2CH_2-$; $-C\equiv C-CH_2-$ and $-CH_2CH_2CH_2-$.

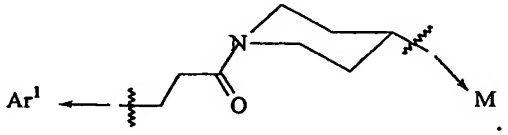
E^M, in the orientation Ar^1-$(L_M\{M\}_p)_q$ or Ar^1-$(L_M\{M'\}_p)_q$, as appropriate, is independently $-Z^M-$, $-C(=Z^M)-$, $-Z^MC(=Z^M)-$, $-C(=Z^M)Z^M-$, $-Z^MC(=Z^M)Z^M-$, $-S(=O)-$, $-Z^MS(=O)-$, $-S(=O)Z^M-$, $-Z^MS(=O)Z^M-$, $-S(=O)_2-$, $-Z^MS(=O)_2-$, $-S(=O)_2Z^M-$, $-Z^MS(=O)_2Z^M-$, where Z^M is independently O, S or $N(R^M)$ and where R^M is independently H, C_{1-8}hydrocarbyl (*e.g.* C_{1-8}alkyl) or C_{1-8}hydrocarbyl substituted with one or more A. Preferably E^M is, in the orientation Ar^1-$(L_M\{M\}_p)_q$ or Ar^1-$(L_M\{M'\}_p)_q$, as appropriate, -O-, -S-, $-C(=O)-$, $-C(=O)O-$, $-C(=S)-$, $-C(=S)O-$, $-OC(=S)-$, $-C(=O)S-$, $-SC(=O)-$, $-S(O)-$, $-S(O)_2-$, $-NR^M-$, $-C(=O)N(R^M)-$, $-C(=S)N(R^M)-$, $-N(R^M)C(=O)-$, $-N(R^M)C(=S)-$, $-S(=O)N(R^M)-$, $-N(R^M)S(=O)-$, $-S(=O)_2N(R^M)-$, $-N(R^M)S(=O)_2-$, $-OC(=O)O-$, $-SC(=O)O-$, $-OC(=O)S-$, $-N(R^M)C(=O)O-$, $-OC(=O)N(R^M)-$, $-N(R^M)C(=O)N(R^M)-$, $-N(R^M)C(=S)N(R^M)-$, $-N(R^M)S(=O)N(R^M)-$ or $-N(R^M)S(=O)_2N(R^M)-$.

Alternative groups E^M to those defined above, in the orientation Ar^1-$(L_M\{M\}_p)_q$ or Ar^1-$(L_M\{M'\}_p)_q$, as appropriate, are $-Z^M-Si(R^M)_2-Z^M-$, $-Si(R^M)_2-Z^M-$ and $-Z^M-Si(R^M)_2-$. The group $-Si(R^M)_2-Z^M-$ is particularly preferred. Z^M is preferably O. R^M is preferably C_{1-8}alkyl, preferably methyl. These groups E^M are particularly preferred in the groups $-(E^M-D^M)_t-$, especially when t=1 and D^M is C_{1-8}alkylene. The following group is especially preferred:



In addition to the above definition of D^M, D^M may also be C_{1-8}heterohydrocarbylene or C_{1-8}heterohydrocarbylene substituted with one or more A. In this embodiment, C_{1-8}cycloheteroalkylene groups are particularly preferred, *e.g.*: 

. Thus, preferred L_M groups $-D^M-E^M-D^M-$ are, in the orientation Ar^1-$(L_M\{M\}_p)_q$ or Ar^1-$(L_M\{M'\}_p)_q$, as appropriate, $-C_{1-8}$alkylene-C(O)-C_{1-8}cycloheteroalkylene (preferably where the hetero atom is N and is bound to the carboxy), especially:



t = 1 or more, *e.g.* from 1 to 50, 1to 40, 1 to 30, 1 to 20 or 1 to 10. Preferably t = 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

Preferably, L_M links one group M (or M') to Ar^1, M (or M') is linked to L_M by a single covalent bond and therefore no additional bonds are required (*e.g.* $L_M\{M\}_1$ may be $-E^M-\{M\}$, $-(D^M)_t-\{M\}$, $-(E^M-D^M)_t-\{M\}$, $-(D^M-E^M)_t-\{M\}$, $-E^M-(D^M-E^M)_t-\{M\}$ or $-D^M-(E^M-D^M)_t-\{M\}$).

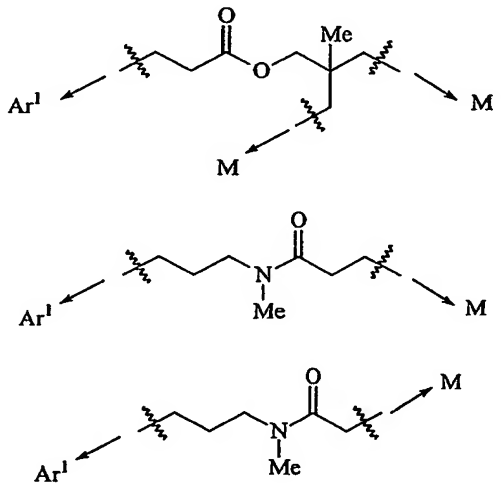
Where L_M includes a group which also falls within the definition of group M, the group M is preferably more reactive than the group included in L_M.

L_M is preferably $-(D^M)_t-$, $-(E^M-D^M)_t-$, or $-D^M-(E^M-D^M)_t-$.

When group L_M is $-(D^M)_t-$, t is preferably 1. D^M is preferably C_{1-8}alkylene, preferably methylene or ethylene.

When group L_M is $-(E^M-D^M)_t-$, or $-D^M-(E^M-D^M)_t-$, E^M is preferably (in the orientation $Ar^1-(L_M\{M\}_p)_q$ or $Ar^1-(L_M\{M'\}_p)_q$, as appropriate), $-C(=O)N(R^M)-$ (*e.g.* -C(=O)NH-) or O (preferably O), and D^M is preferably C_{1-8}alkylene, preferably ethylene, propylene, butylene or pentylene (preferably ethylene or propylene). t is preferably 1. Especially preferred L_M are, in the orientation $Ar^1-(L_M\{M\}_p)_q$ or $Ar^1-(L_M\{M'\}_p)_q$, as appropriate,, $-O-CH_2CH_2CH_2-$ and $-O-CH_2CH_2CH_2CH_2CH_2-$.

Another preferred group $-D^M-(E^M-D^M)_t-$ is where D^M is C_{1-8}alkylene and t is 1. Preferred E^M in this group, in the orientation $Ar^1-(L_M\{M\}_p)_q$ or $Ar^1-(L_M\{M'\}_p)_q$, as appropriate, are $-Z^MC(=Z^M)-$ (especially $-N(R^M)C(=O)-$, *e.g.* -N(Me)C(=O)-) and $-C(=Z^M)Z^M-$ (especially -C(=O)O-). Particularly preferred L_M groups are:



The group $-(E^M-D^M)_t-$ is preferred, a particularly preferred example of which is (in the orientation $Ar^1-(L_M\{M\}_p)_q$ or $Ar^1-(L_M\{M'\}_p)_q$, as appropriate) $-C(=O)NH-CH_2CH_2CH_2-O-CH_2CH_2-O-CH_2CH_2-O-CH_2CH_2CH_2-$.

In an alternative embodiment it is preferred that L_M is a single covalent bond.

When Ar^2 is phenyl, L_M is preferably provided in a position ortho or para to C★. When Ar^2 is other than phenyl, L_M is preferably attached to an atom which bears the charge in at least one of the resonance structures of the ions of formula (I).

Where C★ is a cation, L_M is preferably an electron-donating group. Where C★ is an anion, L_M is preferably an electron-withdrawing group.

Preferred examples of L_M are shown in figure 10A and 10B.

C- - -S_S, S_S- - -Ar^1 and S_S- - -Ar^2 Bonds

C- - -S_S, S_S- - -Ar^1 and S_S- - -Ar^2 comprise a cleavable covalent, ionic, hydrogen, dipole-dipole or van der Waals bond (also known as a dispersion bond or a London forces bond). The covalent, ionic, hydrogen, dipole-dipole or van der Waals bond may be direct between C and S_S, Ar^1 and S_S, or Ar^2 and S_S, or may be provided by one or more binding groups present on C and/or S_S, Ar^1 and/or S_S, or Ar^2 and/or S_S, respectively.

Covalent Bonding

Where the bond is covalent, the bond may be direct (*e.g.* C-S_S, Ar^1-S_S or Ar^2-S_S, respectively) or may be provided by a linker atom or group L^4 (*e.g.* C-L^4-S_S, Ar^1-L^4-S_S or Ar^2-L^4-S_S, respectively).

When L^4 is a linker group, preferred linker groups are -E^4-, -$(D^4)_{t''}$-, -$(E^4$-$D^4)_{t''}$-, -$(D^4$-$E^4)_{t''}$-, -E^4-$(D^4$-$E^4)_{t''}$- or -D^4-$(E^4$-$D^4)_{t''}$-.

D^4 is independently C_{1-8}hydrocarbylene or C_{1-8}hydrocarbylene substituted with one or more A.

E^4 is, in the orientation C-L^4-S_S, independently -Z^4-, -C(=Z^4)-, -Z^4C(=Z^4)-, -C(=Z^4)Z^4-, -Z^4C(=Z^4)Z^4-, -S(=O)-, -Z^4S(=O)-, -S(=O)Z^4-, -Z^4S(=O)Z^4-, -S(=O)$_2$-, -Z^4S(=O)$_2$-, -S(=O)$_2Z^4$-, -Z^4S(=O)$_2Z^4$-, where Z^4 is independently O, S or N(R^4), and where R^4 is independently H, C_{1-8}hydrocarbyl (*e.g.* C_{1-8}alkyl) or C_{1-8}hydrocarbyl substituted with one or more A. Preferably E^4 is, in the orientation C-L^4-S_S, -O-, -S-, -C(=O)-, -C(=O)O-, -C(=S)-, -C(=S)O-, -OC(=S)-, -C(=O)S-, -SC(=O)-, -S(O)-, -S(O)$_2$-, -N(R^4)-, -C(=O)N(R^4)-, -C(=S)N(R^4)-, -N(R^4)C(=O)-, -N(R^4)C(=S)-, -S(=O)N(R^4)-, -N(R^4)S(=O)-, -S(=O)$_2$N(R^4)-, -N(R^4)S(=O)$_2$-, -OC(=O)O-, -SC(=O)O-, -OC(=O)S-, -N(R^4)C(=O)O-, -OC(=O)N(R^4)-, -N(R^4)C(=O)N(R^4)-, -N(R^4)C(=S)N(R^4)-, -N(R^4)S(=O)N(R^4)- or -N(R^4)S(=O)$_2$N(R^4)-.

t'' = 1 or more, *e.g.* from 1 to 50, 1to 40, 1 to 30, 1 to 20 or 1 to 10. Preferably t'' = 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

Where L^4 includes a group which also falls within the definition of group M, the group M is preferably more reactive than the group included in L^5.

L^4 is preferably a linker atom, preferably O or S, particularly O.

When the solid support S_S is gold, L^4 is preferably covalently attached to the S_S by a sulphide or disulphide group.

Ionic Bonding

Where the bond is ionic, the bond is typically direct (*e.g.* C★ S_S★, where S_S★ is a solid support counterion to C★).

Alternatively, it may be provided by binding groups, *e.g.* chelating ligands, present on C or S_S, Ar^1 or S_S, or Ar^2 or S_S, respectively. In the case of C- - -S_S bonds, the chelating ligand is typically only present on S_S and chelates with C★.

Suitable chelating ligands which can bind anions include polyamines and cryptands.

Suitable chelating ligands which can bind cations include polyacidic compounds (*e.g.* EDTA) and crown ethers.

Hydrogen Bonding

Where the bond is a hydrogen bond, the bond is usually provided by binding groups present on C or S_S, Ar^1 or S_S, or Ar^2 or S_S, respectively.

Typically, in order to form the hydrogen bond, one of C or S_S, Ar^1 or S_S, or Ar^2 or S_S, as appropriate, will have a binding group bearing one or more hydroxy, amino or thio hydrogen atoms, and the other of C or S_S, Ar^1 or S_S, or Ar^2 or S_S, respectively, will have a binding group bearing an atom having one or more lone pair of electrons (*e.g.* an oxygen, sulphur or nitrogen atom). Preferably, one of C or S_S, Ar^1 or S_S, or Ar^2 or S_S, as appropriate, will have a binding group comprising biotin, and the other of C or S_S, Ar^1 or S_S, or Ar^2 or S_S, respectively, will have a binding group comprising avidin or streptavidin.

Alternatively, the hydrogen bond may be direct.

Dipole-Dipole Bonding

Where the bond is a dipole-dipole bond, it may be formed between permanent dipoles or between a permanent dipole and an induced dipole.

Typically, in order to form the dipole-dipole bond, one of S_S and the compound of the invention has a permanent dipole and the other of S_S and the compound of the invention has an induced dipole or a permanent dipole, the attraction between the dipoles forming a dipole-dipole bond.

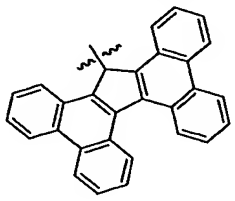
Preferably, S_S comprises binding groups (*e.g.* acid groups, $-(NMe_3)^+$, carboxy, carboxylate, phosphate or sulphate groups) which produce a dipole at the surface of the solid support to bind the compound of the invention.

Van der Waals Bonding

Where the bond is a van der Waals bond, the bonding is usually provided by binding groups present on C or S_S, Ar^1 or S_S, or Ar^2 or S_S, respectively.

Typically, in order to form the van der Waals bond, at least one, but preferably both, of C or S_S, Ar^1 or S_S, or Ar^2 or S_S, as appropriate, will have a hydrocarbyl or heterohydrocarbyl group (usually a large hydrocarbyl group having at least ten carbon atoms up to about 50 carbon atoms), optionally substituted with one or more A. Polyfluorinated hydrocarbyl and heterohydrocarbyl groups are particularly preferred. Typically, the hydrocarbyl or heterohydrocarbyl groups are aryl or heteroaryl groups or groups of the formula $-C(R^6)_2Ar^3$, $-C(R^6)(Ar^3)_2$ or $-C(Ar^3)_3$, where Ar^3 is independently defined the same as Ar^2 and R^6 is H, C_{1-8} hydrocarbyl, C_{1-8} hydrocarbyl substituted by one or more A, C_{1-8} heterohydrocarbyl or C_{1-8} heterohydrocarbyl substituted by one or more A.

A preferred binding group is tetrabenzofullerene (formula X).



(formula X)

Alternatively, the van der Waals bond may be direct.

Bond Cleavage

Preferably, the ions of formula (I) have a pK_{R+} value of at least *zz*, where *zz* is 0 or more (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14). More preferably, *zz* is 1 or more, still more preferably 2 or more, still more preferably 3 or more.

Preferably, the compounds of formula (IIa), (IIb), (IIIa) or (IIIb) or the solid supports of formula (IVai), (IVaii), (IVaiii), (IVbii), (IVbiii), (IVaiv) or (IVbiv) provide ions of formula (I') having a pK_{R+} value of at least *zz*, where *zz* is defined above.

C-X Bonds

The C-X bonds are cleavable by irradiation, electron bombardment, electrospray, fast atom bombardment (FAB), inductively coupled plasma (ICP) or chemical ionisation. Preferably, the C-X bonds are cleavable by irradiation or chemical ionisation.

The term 'irradiation' includes, for example, laser illumination, in particular as used in MALDI mass spectrometry. Laser light of about 340 nm is particularly preferred because it is typically used in MALDI mass spectrometers.

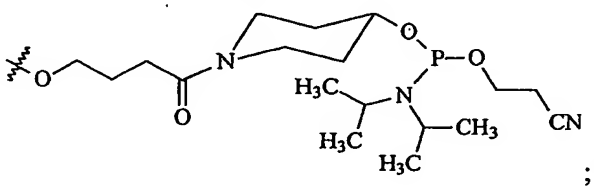
The term 'electron bombardment' includes, for example, bombardment with electrons having energy of about 70 ev.

Chemical ionisation can be effected, for example, by treatment with acid or acidic matrices (*e.g.* acidic matrices used in MALDI analysis).

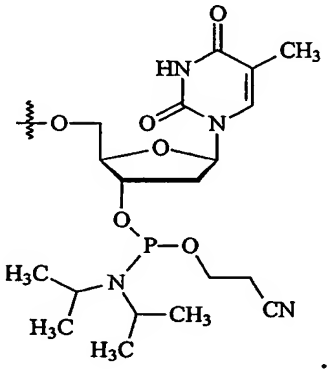
Preferably group X is halogen, hydroxy, C_{1-8}hydrocarbyloxy, C_{1-8}hydrocarbyloxy substituted with one or more A, C_{1-8}heterohydrocarbyloxy, C_{1-8}heterohydrocarbyloxy substituted with one or more A, mesyl, tosyl, pentafluorophenyl, -O-succinimidyl -S-succinimidyl, or phenyloxy substituted with one or more A *e.g.* p-nitrophenyloxy. The groups pentafluorophenyl, -O-succinimidyl, -S-succinimidyl, and p-nitrophenyloxy are particularly preferred.

Particularly preferred groups X are halogen, hydroxy, C_{1-8}hydrocarbyloxy. Especially preferred groups are hydroxy, ethoxy and chloro groups.

Other preferred groups X are alkyl ethers, *e.g.*:

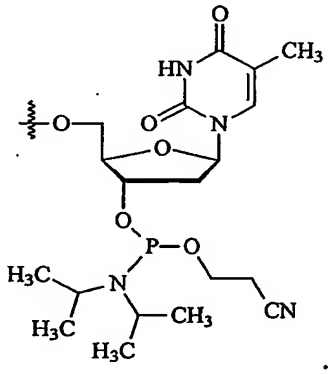


; or

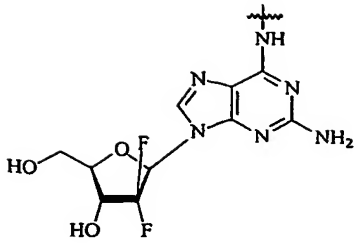


Group X may also be a -Q-oligonucleotide, where Q is O, S or N(R), where R is H, C_{1-8}hydrocarbyl or C_{1-8}hydrocarbyl substituted with one or more A. Q is preferably O.

Group X may also be a nucleoside, preferably where the nucleoside is bound via its 5' end, *e.g.*:



In some embodiments of the invention, where B_P is an antibody (particularly where it is a monoclonal antibody that recognises a tumour-associated antigen), X is not:



or, optionally, X is not any other 2,6-diaminopurine nucleoside prodrug group.

In some embodiments of the invention, X is not H. If X is H, preferably at least one of Ar^1 and Ar^2 is polycyclic, heterocyclic or unsubstituted.

Preferred examples of group X are shown in figure 13.

Ionic C★X★ Bonds

X★ is any counterion for forming salts with compounds of the invention.

X★ includes ions having single charges and multiple charges. Typically ions having multiple charges will be associated with an appropriate number of compounds of formula (IIb), (IIIb), (IVbii), (IVbiii), (IVbiv), (Vbii), (Vbiii) or (Vbiv) in order balance the charge. Ions having multiple charges include doubly charged ions (*e.g.* SO_4^{2-}) and triply charged ions. X★ preferably has a single charge.

The counterion X★ may be dissociated from the derivative of formula (IIb), (IIIb), (IVbii), (IVbiii), (IVbiv), (Vbii), (Vbiii) or (Vbiv) by irradiation, electron bombardment, electrospray, fast atom bombardment (FAB), inductively coupled plasma (ICP) or chemical ionisation. Preferably, the counterion X★ may be dissociated by irradiation.

When X★ is a cation, X★ is preferably H^+.

When X★ is an anion, X★ is preferably, BF_6^- or ClO_4^-.

It is preferred that X★ is an anion.

Preferred examples of group X★ are shown in figure 13.

C---S_S, S_S---Ar^1 or S_S---Ar^2

The C- - -S_S, S_S- - -Ar^1 or S_S- - -Ar^2 bonds are cleavable by irradiation, electron bombardment, electrospray, fast atom bombardment (FAB), inductively coupled plasma (ICP) or chemical ionisation. Preferably, the C---S_S, S_S---Ar^1 or S_S---Ar^2 bonds are cleavable by irradiation or chemical ionisation.

Where appropriate, the C- - -S_S, S_S- - -Ar^1 or S_S- - -Ar^2 bonds may be cleaved simultaneously or sequentially with the cleaving of the C-X bond or the dissociation of X★, as appropriate, by selection of suitable cleaving/dissociating conditions.

In one embodiment of the invention, the C- - -S_S bond in the solid support of formula (Vai) may be cleaved in sub-steps of step (iia) so that in a first sub-step a residue X (where X is the leaving group

defined above) is provided and in a second subsequent sub-step the C-X bond is cleaved thereby forming the ion of formula (I). If desired, the second sub-step may be carried out substantially (*e.g.* seconds, minutes, hours or even days) after the first sub-step.

Ar^1 and Ar^2

Ar^2

Ar^2 is independently an aromatic group or an aromatic group substituted with one or more A and is preferably independently cyclopropyl, cyclopropyl substituted with one or more A, aryl, aryl substituted with one or more A, heteroaryl, or heteroaryl substituted with one or more A.

Where aryl or substituted aryl, Ar^2 is preferably C_{6-30} aryl or substituted C_{6-30} aryl. Where heteroaryl or substituted heteroaryl, Ar^2 is preferably C_{6-30} heteroaryl or substituted C_{6-30} heteroaryl.

Examples of aryl and heteroaryl are monocyclic aromatic groups (*e.g.* phenyl or pyridyl), fused polycyclic aromatic groups (*e.g.* napthyl, such as 1-napthyl or 2-napthyl) and unfused polycyclic aromatic groups (*e.g.* monocyclic or fused polycyclic aromatic groups linked by a single bond, a double bond, or by a $-(CH=CH)_r-$ linking group, where r is one or more (*e.g.* 1, 2, 3, 4 or 5).

Other examples of aryl groups are monovalent derivatives of aceanthrylene, acenaphthylene, acephenanthrylene, anthracene, azulene, chrysene, coronene, fluoranthene, fluorene, *as*-indacene, *s*-indacene, indene, naphthalene, ovalene, perylene, phenalene, phenanthrene, picene, pleiadene, pyrene, pyranthrene and rubicene, which groups may be optionally substituted by one or more A.

Other examples of heteroaryl groups are monovalent derivatives of acridine, carbazole, β-carboline, chromene, cinnoline, furan, imidazole, indazole, indole, indolizine, isobenzofuran, isochromene, isoindole, isoquinoline, isothiazole, isoxazole, naphthyridine, perimidine, phenanthridine, phenanthroline, phenazine, phthalazine, purine, pyran, pyrazine, pyrazole, pyridazine, pyridine, pyrimidine, pyrrole, pyrrolizine, quinazoline, quinoline, quinolizine, quinoxaline, thiophene and xanthene, which groups may be optionally substituted by one or more A. Preferred heteroaryl groups are five- and six-membered monovalent derivatives, such as the monovalent derivatives of furan, imidazole, isothiazole, isoxazole, pyran, pyrazine, pyrazole, pyridazine, pyridine, pyrimidine, pyrrole, pyrrolizine and thiophene. The five-membered monovalent derivatives are particularly preferred, *i.e.* the monovalent derivatives of furan, imidazole, isothiazole, isoxazole, pyrazole, pyrrole and thiophene. The heteroaryl groups may be attached to the remainder of the compound by any carbon or hetero (*e.g.* nitrogen) atom.

Ar^2 is preferably C_{6-30}aryl substituted by one or more A, preferably phenyl or napthyl (*e.g.* 1-napthyl or 2-napthyl, especially 2-napthyl) substituted by one or more A, more preferably phenyl substituted by one or more A. When Ar^2 is phenyl, A is preferably provided in a position ortho or para to C★. When Ar^2 is other than phenyl, A is preferably attached to an atom which bears the charge in at least one of the resonance structures of the ions of formula (I).

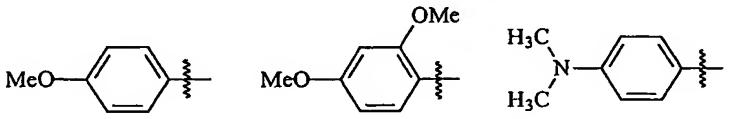
Fused polycyclic aromatic groups, optionally substituted with one or more A, are particularly preferred.

A particularly preferred Ar^2 is unsubstituted pyrenyl or pyrenyl substituted with one or more A. Unsubstituted pyrenyl is preferred. The pyrenyl group may be 1-pyrenyl, 2-pyrenyl or 4-pyrenyl.

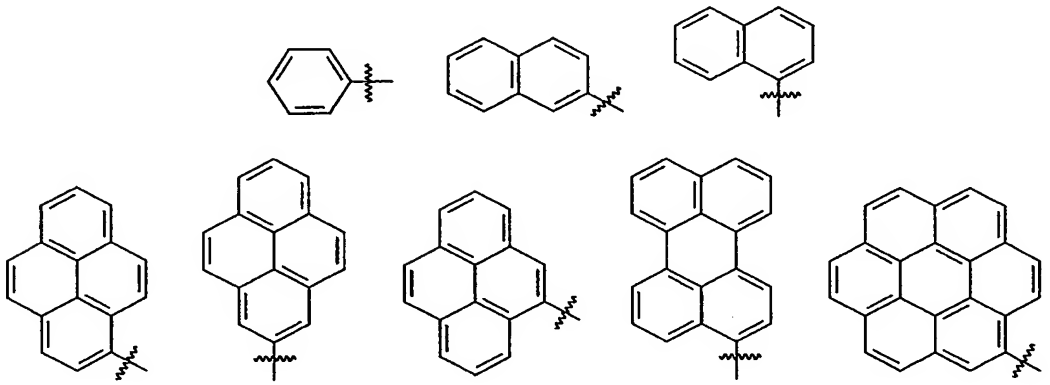
Preferred heteroaryl Ar^2 groups, whether substituted or unsubstituted, are pyridyl, pyrrolyl, thienyl and furyl, especially thienyl.

A preferred Ar^2 group is thiophenyl or thiophenyl substituted with one or more A. Unsubstituted thiophenyl is preferred. Examples of thiophenyl are thiophen-2-yl and thiophen-3-yl, with thiophen-2-yl being especially preferred.

When substituted, Ar^2 is preferably substituted by 1, 2 or 3 A. Ar^2 is preferably:



When unsubstituted, Ar^2 is preferably:



In another preferred embodiment, Ar^2 is cyclopropyl or cyclopropyl substituted with one or more A. Unsubstituted cyclopropyl is preferred. One or more, preferably one, of Ar^2 may be cyclopropyl.

Preferred examples of group Ar^2 are shown in figures 12A and 12B.

Ar^1

Ar^1 is independently an aromatic group or an aromatic group substituted with one or more A. The definition of Ar^1 is the same as Ar^2 (as defined above), except that the valency of the group Ar^1 is adapted to accommodate the q instances of the linker L_M. Preferred Ar^2 groups are also preferred Ar^1 groups, (as defined above), except that the valency of the group Ar^1 is adapted to accommodate the q instances of the linker L_M.

When q = 1, Ar^1 is a divalent radical and is preferably independently cyclopropylene, cyclopropylene substituted with one or more A, arylene, arylene substituted with one or more A, heteroarylene, or heteroarylene substituted with one or more A.

Where arylene or substituted arylene, Ar^1 is preferably C_{6-30} arylene or substituted C_{6-30} arylene. Where heteroarylene or substituted heteroarylene, Ar^1 is preferably C_{6-30} heteroarylene or substituted C_{6-30} heteroarylene.

Examples of arylene and heteroarylene are monocyclic aromatic groups (*e.g.* phenylene or pyridylene), fused polycyclic aromatic groups (*e.g.* napthylene) and unfused polycyclic aromatic groups (*e.g.* monocyclic or fused polycyclic aromatic groups linked by a single bond, a double bond, or by a $-(CH=CH)_r-$ linking group, where r is one or more (*e.g.* 1, 2, 3, 4 or 5).

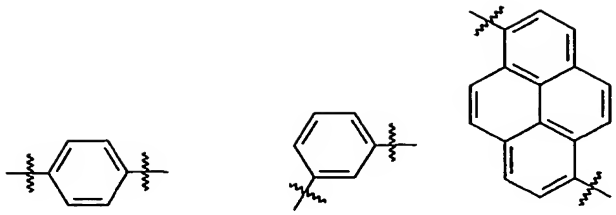
Other examples of arylene groups are polyvalent derivatives (where the valency is adapted to accommodate the q instances of the linker L_M) of aceanthrylene, acenaphthylene, acephenanthrylene, anthracene, azulene, chrysene, coronene, fluoranthene, fluorene, *as*-indacene, *s*-indacene, indene, naphthalene, ovalene, perylene, phenalene, phenanthrene, picene, pleiadene, pyrene, pyranthrene and rubicene, which groups may be optionally substituted by one or more A.

Other examples of heteroarylene groups are polyvalent derivatives (where the valency is adapted to accommodate the q instances of the linker L_M) of acridine, carbazole, β-carboline, chromene, cinnoline, furan, imidazole, indazole, indole, indolizine, isobenzofuran, isochromene, isoindole, isoquinoline, isothiazole, isoxazole, naphthyridine, perimidine, phenanthridine, phenanthroline, phenazine, phthalazine, purine, pyran, pyrazine, pyrazole, pyridazine, pyridine, pyrimidine, pyrrole, pyrrolizine, quinazoline, quinoline, quinolizine, quinoxaline, thiophene and xanthene, which groups may be optionally substituted by one or more A. Preferred heteroaryl groups are five- and six-membered polyvalent derivatives, such as the polyvalent derivatives of furan, imidazole, isothiazole, isoxazole, pyran, pyrazine, pyrazole, pyridazine, pyridine, pyrimidine, pyrrole, pyrrolizine and thiophene. The five-membered polyvalent derivatives are particularly preferred, *i.e.* the polyvalent derivatives of furan, imidazole, isothiazole, isoxazole, pyrazole, pyrrole and thiophene. The heteroaryl groups may be attached to the remainder of the compound by any carbon or hetero (*e.g.* nitrogen) atom.

Ar^1 is preferably C_{6-30}arylene substituted by one or more A, preferably phenylene or napthylene substituted by one or more A, more preferably phenylene substituted by one or more A. When Ar^1 is phenylene, A is preferably provided in a position ortho or para to C★. When Ar^1 is other than phenylene, A is preferably attached to an atom which bears the charge in at least one of the resonance structures of the ions of formula (I).

When substituted, Ar^1 is preferably substituted by 1, 2 or 3 A.

When unsubstituted, preferred Ar^1 are:



Preferred examples of group Ar^1 are shown in figures 12A and 12B.

Combinations of Ar

Optionally two or three of the groups Ar^1 and Ar^2 are linked together by one or more L^5, where L^5 is independently a single bond or a linker atom or group; and/or two or three of the groups Ar^1 and Ar^2 together form an aromatic group or an aromatic group substituted with one or more A.

When L^5 is a linker group, preferred linker groups are $-E^5-$, $-(D^5)_{t'}-$, $-(E^5-D^5)_{t'}-$, $-(D^5-E^5)_{t'}-$, $-E^5-(D^5-E^5)_{t'}-$ or $-D^5-(E^5-D^5)_{t'}-$.

D^5 is independently C_{1-8}hydrocarbylene or C_{1-8}hydrocarbylene substituted with one or more A.

E^5 is independently $-Z^5-$, $-C(=Z^5)-$, $-Z^5C(=Z^5)-$, $-C(=Z^5)Z^5-$, $-Z^5C(=Z^5)Z^5-$, $-S(=O)-$, $-Z^5S(=O)-$, $-S(=O)Z^5-$, $-Z^5S(=O)Z^5-$, $-S(=O)_2-$, $-Z^5S(=O)_2-$, $-S(=O)_2Z^5-$, $-Z^5S(=O)_2Z^5-$, where Z^5 is independently O, S or $N(R^5)$ and where R^5 is independently H, C_{1-8}hydrocarbyl or C_{1-8}hydrocarbyl substituted with one or more A. Preferably E^5 is -O-, -S-, $-C(=O)-$, $-C(=O)O-$, $-C(=S)-$, $-C(=S)O-$, $-OC(=S)-$, $-C(=O)S-$, $-SC(=O)-$, $-S(O)-$, $-S(O)_2-$, $-N(R^5)-$, $-C(=O)N(R^5)-$, $-C(=S)N(R^5)-$, $-N(R^5)C(=O)-$, $-N(R^5)C(=S)-$, $-S(=O)N(R^5)-$, $-N(R^5)S(=O)-$, $-S(=O)_2N(R^5)-$, $-N(R^5)S(=O)_2-$, $-OC(=O)O-$, $-SC(=O)O-$, $-OC(=O)S-$, $-N(R^5)C(=O)O-$, $-OC(=O)N(R^5)-$, $-N(R^5)C(=O)N(R^5)-$, $-N(R^5)C(=S)N(R^5)-$, $-N(R^5)S(=O)N(R^5)-$ or $-N(R^5)S(=O)_2N(R^5)-$.

t' = 1 or more, *e.g.* from 1 to 50, 1to 40, 1 to 30, 1 to 20 or 1 to 10. Preferably t' = 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. Most preferably t'=1.

Where L^5 includes an atom or group which also falls within the definition of group M, the group M is preferably more reactive than the group included in L^5.

L^5 is preferably a linker atom, preferably O or S, particularly O.

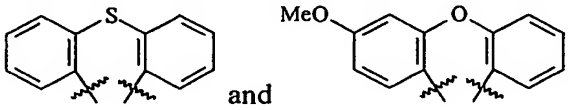
When L^5 is a linker group, a preferred L^5 is $-N(R^5)-$.

In another embodiment in which L^5 is a linker group, L^5 is -S(=O)-.

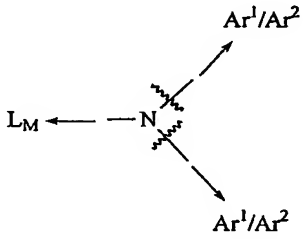
When two of the groups Ar^1 and Ar^2 are linked together by one or more (*e.g.* 2, 3 or 4) L^5, they are preferably linked together by one L^5, preferably O.

Preferred combinations of Ar are two Ar^2 (*e.g.* two Ar^2 phenyl groups) linked together by one L^5 (e.g. O or S).

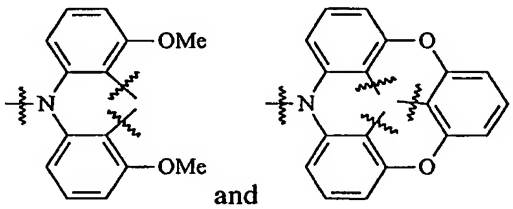
Particularly preferred combinations of Ar are two Ar^2 phenyl groups, optionally substituted by one or more A (preferably unsubstituted), linked together by one L^5 (*e.g.* O or S), where is L^5 is ortho to C★ with respect to both phenyl groups. Especially preferred combinations of two Ar^2 groups are:



In another embodiment, at least one L_M is linked to an atom or group L^5. In this embodiment, the preferred L^5 mentioned above are, where appropriate, modified to remove substituents R^5 in order to accommodate L_M, *e.g.* the R^5 substituent of the group $-N(R^5)-$ is replaced by L_M. In this embodiment, the L^5 group to which L_M is bound is preferably:



Preferred combinations of Ar^1 and/or Ar^2 in this embodiment are:



When two or three of the groups Ar^1 and Ar^2 together form an aromatic group or an aromatic group substituted with one or more A, the aromatic group may be a carbocyclic aromatic group or a carbocyclic aromatic group in which one or more carbon atoms are each replaced by a hetero atom. Typically, in an aromatic group in which one or more carbon atoms are each replaced by a hetero atom, up to three carbons are so replaced, preferably up to two carbon atoms, more preferably one carbon atom.

Preferred hetero atoms are O, Se, S or N, more preferably O, S or N.

When two or three of the groups Ar^1 and Ar^2 together form an aromatic group or an aromatic group substituted with one or more A, preferred aromatic groups are C_{8-50} aromatic groups.

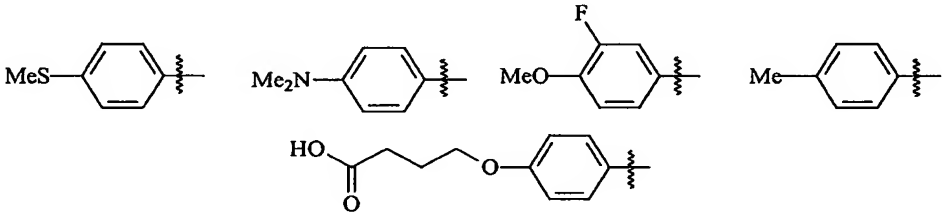
The aromatic groups may be monocyclic aromatic groups (*e.g.* radicals of suitable valency derived from benzene), fused polycyclic aromatic groups (*e.g.* radicals of suitable valency derived from napthalene) and unfused polycyclic aromatic groups (*e.g.* monocyclic or fused polycyclic aromatic groups linked by a single bond, a double bond, or by a $-(CH=CH)_r-$ linking group, where r is one or more (*e.g.* 1, 2, 3, 4 or 5).

When two or three of the groups Ar^1 and Ar^2 together form a carbopolycyclic fused ring aromatic group, preferred groups are radicals of suitable valency obtained from napthalene, anthracene or phenanthracene, chrysene, aceanthrylene, acenaphthylene, acephenanthrylene, azulene, fluoranthene, fluorene, *as*-indacene, *s*-indacene, indene, phenalene, and pleiadene.

When two or three of the groups Ar^1 and Ar^2 together form a carbopolycyclic fused ring aromatic group in which one or more carbon atoms are each replaced by a hetero atom, preferred groups are radicals of suitable polyvalency obtained from acridine, carbazole, β-carboline, chromene, cinnoline, indole, indolizine, isobenzofuran, isochromene, isoindole, isoquinoline, naphthyridine, perimidine, phenanthridine, phenanthroline, phenazine, phthalazine, pteridine, purine, pyrrolizine, quinazoline, quinoline, quinolizine and quinoxaline.

Substitution of Ar^1 and Ar^2 – Anions and Cations

When C★ is a cation, A is preferably an electron-donating group, including $-R^1$ or $-Z^1R^1$, where R^1 and Z^1 are defined below. Preferably, R^1 is C_{1-8}hydrocarbyl, more preferably C_{1-8}alkyl, especially methyl. Z^1 is preferably O, S or NR^1. R^1 may be substituted with one or more S_{ub}^2, but is preferably unsubstituted. When C★ is a cation, A is preferably –OMe, -SMe, $-N(Me)_2$ or Me. When C★ is a cation, A, when an electron-donating group, is preferably provided (especially in relation to Ar^1 or Ar^2 being phenyl) in a position ortho or para to C★, preferably para. Furthermore, when C★ is a cation, A, when an electron-withdrawing group (*e.g.* F), is preferably provided (especially in relation to Ar^1 or Ar^2 being phenyl) in a position meta to C★. Thus, preferred groups Ar^1 and Ar^2 are as follows:



When C★ is an anion, A is preferably an electron-withdrawing group, including halogen, trihalomethyl, $-NO_2$, -CN, $-N^+(R^1)_2O^-$, $-CO_2H$, $-CO_2R^1$, $-SO_3H$, $-SOR^1$, $-SO_2R^1$, $-SO_3R^1$, $-OC(=O)OR^1$, $-C(=O)H$, $-C(=O)R^1$, $-OC(=O)R^1$, $-C(=O)NH_2$, $-C(=O)NR^1{}_2$, $-N(R^1)C(=O)OR^1$, $-N(R^1)C(=O)NR^1{}_2$, $-OC(=O)NR^1{}_2$, $-N(R^1)C(=O)R^1$, $-C(=S)NR^1{}_2$, $-NR^1C(=S)R^1$, $-SO_2NR^1{}_2$, $-NR^1SO_2R^1$, $-N(R^1)C(=S)NR^1{}_2$, or $-N(R^1)SO_2NR^1{}_2$, where R^1 is defined below. When C★ is an anion, A, when an electron-withdrawing group, is preferably provided (especially in relation to Ar^1 or Ar^2 being phenyl) in a position ortho or para to C★, preferably para. Furthermore, when C★ is an anion, A, when an electron-donating group, is preferably provided (especially in relation to Ar^1 or Ar^2 being phenyl) in a position meta to C★.

The group A may also comprise one or more isotopes of the atoms making up group A (*e.g.* example 60), thus, as discussed in more detail below, allowing the masses of the compounds of the invention

to be varied. Preferred isotopes are ^{13}C, ^{18}O and ^{2}H. When providing a series of compounds which differ only in their masses, ^{13}C and ^{18}O are particularly preferred as ^{2}H atoms may cause a substantial change in the chemical properties of the compound due to the kinetic isotope effect.

Solid Supports

'Solid supports' for use with the invention include polymer beads, metals, resins, columns, surfaces (including porous surfaces) and plates (*e.g.* mass-spectrometry plates).

The solid support is preferably one suitable for use in a mass spectrometer, such that the invention can be conveniently accommodated into existing MS apparatus. Ionisation plates from mass spectrometers are thus preferred solid supports, *e.g.* gold, glass-coated or plastic-coated plates. Solid gold supports are particularly preferred.

Resins or columns, such as those used in affinity chromatography and the like, are particularly useful for receiving solutions of biopolymers (purified or mixtures). For example, a cellular lysate could be passed through such a column of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv) followed by cleavage of the support to leave compounds of formula (I).

Solid supports of formulae (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv) will generally present exposed groups M capable of reacting with a biopolymer, B_P. For MS analysis, ions preferably have a predictable mass to charge (m/e) ratio. If a biopolymer reacts with more than one M group, however, then it will carry more than one positive charge once ionised, and its m/e ratio will decrease. Advantageously, therefore, the groups M are arranged such that any biopolymer molecule will covalently link with only a single group M. Consequently, each biopolymer will, on ionisation, carry a single positive charge and thus have a predictable mass to charge ratio.

Typically, the surface density of the solid supports of (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv) will be provided so that a biopolymer molecule can only covalently link with one group M and thus to prevent the formation of multiply derivatised biopolymers.

Varying the mass of compounds of the invention

Within the general formulae (I), (IIa), (IIb), (IIIa), (IIIb), (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii), (IVbiv), (Vai), (Vaii), (Vaiii), (Vaiv), (Vbii), (Vbiii) and (Vbiv), there is much scope for variation. There is thus much scope of variation in the mass of these compounds. In some embodiments of the invention, it is preferred to use a series of two or more (*e.g.* 2, 3, 4, 5, 6 or more) compounds with different and defined molecular masses.

The masses of the compounds of the invention can be varied via L_M, Ar^1 and/or Ar^2. Preferably, the masses of the compounds of the invention are varied by varying A on the groups Ar^1 and/or Ar^2.

In this aspect of invention, compounds of the invention advantageously comprise one or more of F or I as substituents A of the groups Ar^1, Ar^2 or Ar^3. F and I each only have one naturally occurring isotope, ^{19}F and ^{127}I respectively, and thus by varying the number of F and I atoms present in the

structure of the compounds, can provide a series of molecular mass labels having substantially identical shaped peaks on a mass spectrum.

Compounds of the invention may also include one or more ^2H atoms, preferably as a substituent A or a part thereof of the groups L_M, Ar^1, Ar^2 or Ar^3 (in particular L_M), in order to vary the masses of the compounds of the invention. The compounds of the invention may include isotopes of ^{13}C and ^{18}O, prefererably as a substituent A or a part thereof of the groups L_M, Ar^1, Ar^2 or Ar^3 (in particular Ar^1, Ar^2 or Ar^3), in order to vary the masses of the compounds of the invention. Compounds comprising ^2H, ^{13}C and ^{18}O may also be used to provide a series of molecular mass labels having substantially identical shaped peaks on a mass spectrum, by varing the number of ^2H, ^{13}C and ^{18}O atoms present in the structure of the compounds. When providing a series of compounds which differ only in their masses, ^{13}C and ^{18}O are particularly preferred as ^2H atoms may cause a substantial change in the chemical properties of the compound due to the kinetic isotope effect.

In order to increase the molecular mass of the compounds of the invention and to increase the number of available sites for substitution by A, especially F and I, one or more of Ar^1 and Ar^2 may be substituted by one or more dendrimer radicals of appropriate valency, either as substituent A or group L_M.

Preferred dendrimer radicals are the radicals obtained from the dendrimers of US 6,455,071 and PAMAM dendrimers.

The compounds of the invention may advantageously be used in the method of analysing a biopolymer disclosed herein, in particular in a method for following a reaction involving a biopolymer, B_P, since the abundance of a species of may be determined by mass spectrometry by measuring the intensity of the relevant peak in an obtained mass spectrum.

Specifically, there is provided a method for analysing biopolymer B_P, comprising the steps of:

(i) reacting a first sample comprising biopolymer B_P with a compound of formula (IIa) or (IIb) or a solid support of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv) at a time t_1;

(ii) reacting a second sample comprising biopolymer B_P with a compound of formula (IIa) or (IIb) or a solid support of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv) at a later time t_2;

(iii) preparing and analysing cations of formula (I) from the first and second samples; and

(iv) comparing the results of the analysis from step (iii).

If levels of the biopolymer B_P decrease between times t_1 and t_2 then there will be a decrease in detected ion; if levels of the biopolymer B_P increase between times t_1 and t_2 then there will be an increase in detected ion. The effects of stimuli on transcription and/or translation can therefore be monitored.

Advantageously, different compounds of formula (IIa) or (IIb) or different solid supports of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv) are used at different times in order to facilitate simultaneous and parallel analysis of the first and second samples. For example, if the two compounds used at times t_1 and t_2 differ only by a ^1H to ^{19}F substitution then the relative abundance of B_P at the two times can be determined by comparing peaks separated by 18 units.

Advantageously, the reaction of the biopolymer with the compound of formula (IIa) or (IIb) or the solid support of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv) will fix the biopolymer to prevent it reacting further and the steps of providing and analysing the cations may be carried out at a later convenient time. Alternatively, if the reaction of the biopolymer with the compound of formula (IIa) or (IIb) or the solid support of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv) does not quench the reaction of the biopolymer being followed, a cation of formula (I) from the reaction product of step (i) or step (v) should be obtained as soon as possible after reaction of the biopolymer with the compound of formula (IIa) or (IIb) or the solid support of formula (IVai), (IVaii), (IVaiii), (IVaiv), (IVbii), (IVbiii) or (IVbiv).

Compounds of Formulae (IIa) and (IIb)

The compounds of formulae (IIa) or (IIb) are available commercially or may be synthesised by known techniques.

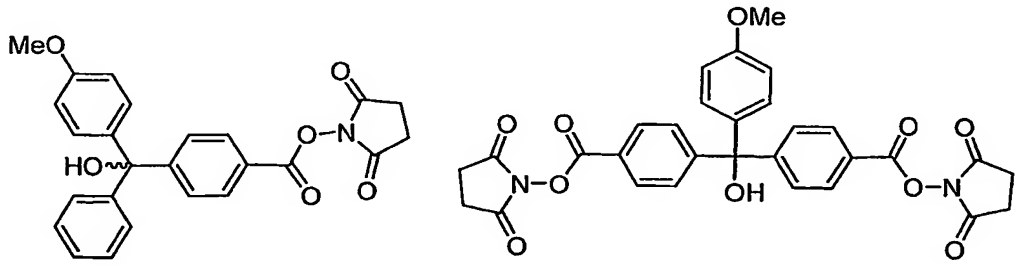
Commercially available compounds of formulae (IIa) or (IIb) are disclosed, for example in the Molecular Probes Catalogue, 2002. Commercially available trityls, and derivatives and analogues thereof, may also be derivatised with the groups $(L_M\{M\}_p)_q$ by known techniques.

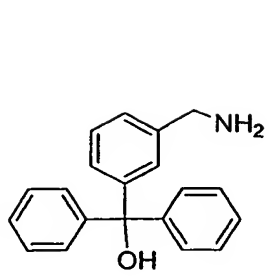
Methods for synthesis of compounds of formula (IIa) or (IIb) useful in the present invention are described in Chem. Soc. Rev. (2003) 32, p. 3-13, scheme 2 and "1. introduction", last two paragraphs. Groups $(L_M\{M\}_p)_q$ are usually introduced into the intermediates and the compounds are then assembled using the appropriate pathways. Alternatively, the groups $(L_M\text{-}\{M\}_p)_q$ may be added after assembly of the aromatic groups and α-carbon of the compounds.

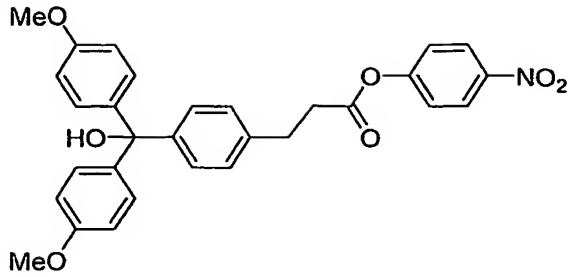
Methods for synthesis of compounds of formulae (IIa) or (IIb) are also described in WO99/60007.

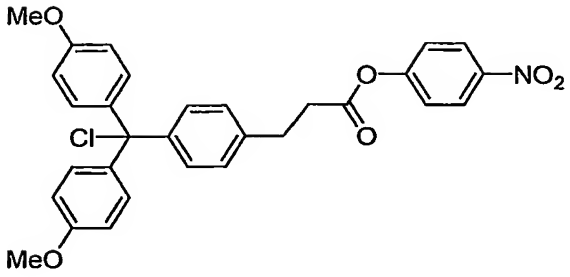
Further methods for synthesising the compounds of formulae (IIa) or (IIb) are described in European patent application 04 104 605.3.

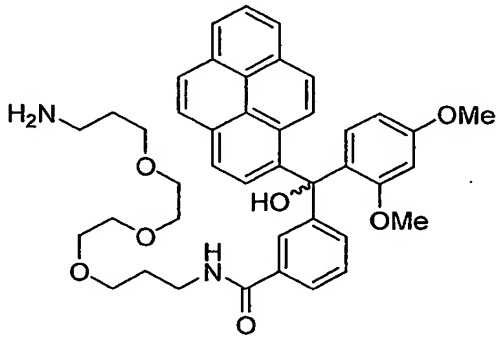
Preferred compounds of formula (IIa), (IIb) and (IVai) are:

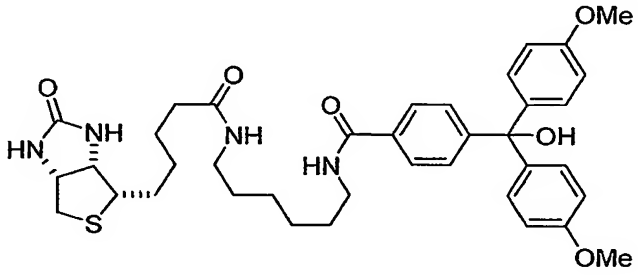


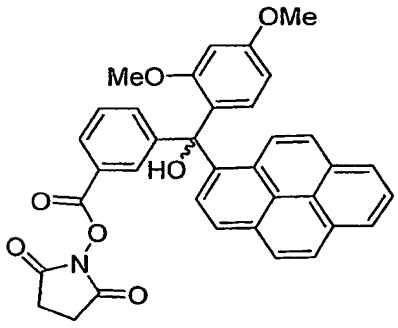

NH2
OH

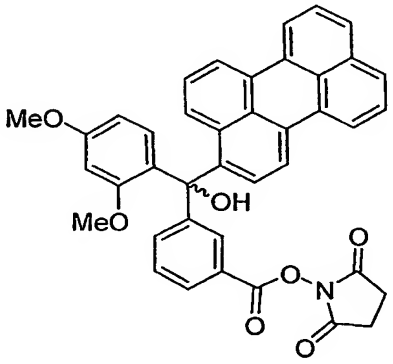

MeO
HO
O
NO2
MeO

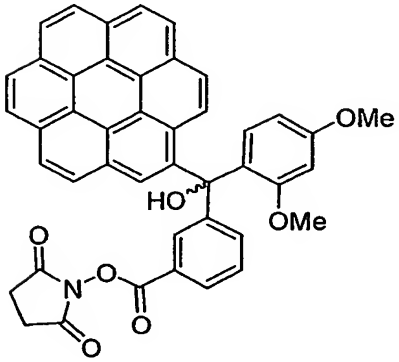

MeO
Cl
O
NO2
MeO

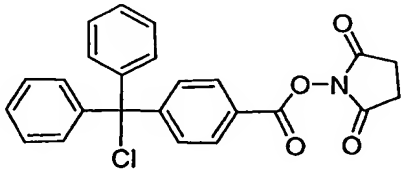

H2N
OMe
HO
OMe
H
N
O

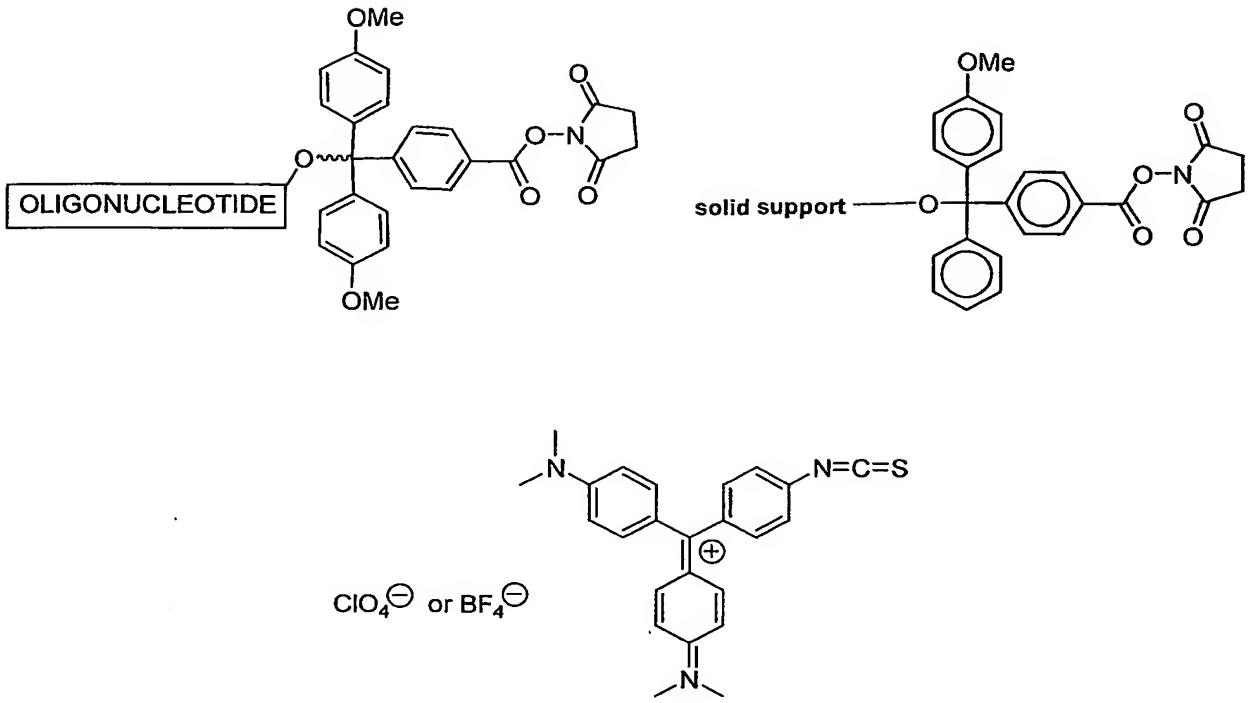

OMe
O
NH
HN
NH
HN
OH
S
OMe


OMe
MeO
HO
O
O
N
O


MeO
MeO
OH
O
O
N
O


OMe
HO
OMe
O
N
O
O


O
O
N
O
O
Cl



Chemical Groups

The ions of the invention are stabilised by the resonance effect of the aromatic groups Ar^1 and Ar^2. The term 'C★ is a carbon atom bearing a single positive charge or a single negative charge' therefore not only includes structures having the charge localised on the carbon atom but also resonance structures in which the charge is delocalised from the carbon atom.

The term 'linker atom or group' includes any divalent atom or divalent group.

The term 'aromatic group' includes quasi and/or pseudo-aromatic groups, *e.g.* cyclopropyl and cyclopropylene groups.

The term 'halogen' includes fluorine, chlorine, bromine and iodine.

The term 'hydrocarbyl' includes linear, branched or cyclic monovalent groups consisting of carbon and hydrogen. Hydrocarbyl groups thus include alkyl, alkenyl and alkynyl groups, cycloalkyl (including polycycloalkyl), cycloalkenyl and aryl groups and combinations thereof, *e.g.* alkylcycloalkyl, alkylpolycycloalkyl, alkylaryl, alkenylaryl, cycloalkylaryl, cycloalkenylaryl, cycloalkylalkyl, polycycloalkylalkyl, arylalkyl, arylalkenyl, arylcycloalkyl and arylcycloalkenyl groups. Preferred hydrocarbyl are C_{1-14} hydrocarbyl, more preferably C_{1-8} hydrocarbyl.

Unless indicated explicitly otherwise, where combinations of groups are referred to herein as one moiety, *e.g.* arylalkyl, the last mentioned group contains the atom by which the moiety is attached to the rest of the molecule.

The term 'hydrocarbylene' includes linear, branched or cyclic divalent groups consisting of carbon and hydrogen formally made by the removal of two hydrogen atoms from the same or different (preferably different) skeletal atoms of the group. Hydrocarbylene groups thus include alkylene,

alkenylene and alkynylene groups, cycloalkylene (including polycycloalkylene), cycloalkenylene and arylene groups and combinations thereof, *e.g.* alkylenecycloalkylene, alkylenepolycycloalkylene, alkylenearylene, alkenylenearylene, cycloalkylenealkylene, polycycloalkylenealkylene, arylenealkylene and arylenealkenylene groups. Preferred hydrocarbylene are C_{1-14} hydrocarbylene, more preferably C_{1-8} hydrocarbylene.

The term 'hydrocarbyloxy' means hydrocarbyl-O-.

The terms 'alkyl', 'alkylene', 'alkenyl', 'alkenylene', 'alkynyl', or 'alkynylene' are used herein to refer to both straight, cyclic and branched chain forms. Cyclic groups include C_{3-8} groups, preferably C_{5-8} groups.

The term 'alkyl' includes monovalent saturated hydrocarbyl groups. Preferred alkyl are C_{1-8}, more preferably C_{1-4} alkyl such as methyl, ethyl, n-propyl, i-propyl or t-butyl groups.

Preferred cycloalkyl are C_{5-8} cycloalkyl.

The term 'alkoxy' means alkyl-O-.

The term 'alkenyl' includes monovalent hydrocarbyl groups having at least one carbon-carbon double bond and preferably no carbon-carbon triple bonds. Preferred alkenyl are C_{2-4} alkenyl.

The term 'alkynyl' includes monovalent hydrocarbyl groups having at least one carbon-carbon triple bond and preferably no carbon-carbon double bonds. Preferred alkynyl are C_{2-4} alkynyl.

The term 'aryl' includes monovalent aromatic groups, such as phenyl or naphthyl. In general, the aryl groups may be monocyclic or polycyclic fused ring aromatic groups. Preferred aryl are C_6-C_{14}aryl.

Other examples of aryl groups are monovalent derivatives of aceanthrylene, acenaphthylene, acephenanthrylene, anthracene, azulene, chrysene, coronene, fluoranthene, fluorene, *as*-indacene, *s*-indacene, indene, naphthalene, ovalene, perylene, phenalene, phenanthrene, picene, pleiadene, pyrene, pyranthrene and rubicene.

The term 'alkylene' includes divalent saturated hydrocarbylene groups. Preferred alkylene are C_{1-4} alkylene such as methylene, ethylene, n-propylene, i-propylene or t-butylene groups.

Preferred cycloalkylene are C_{5-8} cycloalkylene.

The term 'alkenylene' includes divalent hydrocarbylene groups having at least one carbon-carbon double bond and preferably no carbon-carbon triple bonds. Preferred alkenylene are C_{2-4} alkenylene.

The term 'alkynylene' includes divalent hydrocarbylene groups having at least one carbon-carbon triple bond and preferably no carbon-carbon double bonds. Preferred alkynylene are C_{2-4} alkynylene.

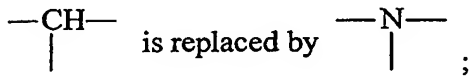
The term 'arylene' includes divalent aromatic groups, such phenylene or naphthylene. In general, the arylene groups may be monocyclic or polycyclic fused ring aromatic groups. Preferred arylene are C_6-C_{14}arylene.

Other examples of arylene groups are divalent derivatives of aceanthrylene, acenaphthylene, acephenanthrylene, anthracene, azulene, chrysene, coronene, fluoranthene, fluorene, *as*-indacene, *s*-indacene, indene, naphthalene, ovalene, perylene, phenalene, phenanthrene, picene, pleiadene, pyrene, pyranthrene and rubicene.

The term 'heterohydrocarbyl' includes hydrocarbyl groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N. Heterohydrocarbyl groups thus include heteroalkyl, heteroalkenyl and heteroalkynyl groups, cycloheteroalkyl (including polycycloheteroalkyl), cycloheteroalkenyl and heteroaryl groups and combinations thereof, *e.g.* heteroalkylcycloalkyl, alkylcycloheteroalkyl, heteroalkylpolycycloalkyl, alkylpolycycloheteroalkyl, heteroalkylaryl, alkylheteroaryl, heteroalkenylaryl, alkenylheteroaryl, cycloheteroalkylaryl, cycloalkylheteroaryl, heterocycloalkenylaryl, cycloalkenylheteroaryl, cycloalkylheteroalkyl, cycloheteroalkylalkyl, polycycloalkylheteroalkyl, polycycloheteroalkylalkyl, arylheteroalkyl, heteroarylalkyl, arylheteroalkenyl, heteroarylalkenyl, arylcycloheteroalkyl, heteroarylcycloalkyl, arylheterocycloalkenyl and heteroarylcycloalkenyl groups. The heterohydrocarbyl groups may be attached to the remainder of the compound by any carbon or hetero (*e.g.* nitrogen) atom.

The term 'heterohydrocarbylene' includes hydrocarbylene groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N. Heterohydrocarbylene groups thus include heteroalkylene, heteroalkenylene and heteroalkynylene groups, cycloheteroalkylene (including polycycloheteroalkylene), cycloheteroalkenylene and heteroarylene groups and combinations thereof, *e.g.* heteroalkylenecycloalkylene, alkylenecycloheteroalkylene, heteroalkylenepolycycloalkylene, alkylenepolycycloheteroalkylene, heteroalkylenearylene, alkyleneheteroarylene, heteroalkenylenearylene, alkenyleneheteroarylene, cycloalkyleneheteroalkylene, cycloheteroalkylenealkylene, polycycloalkyleneheteroalkylene, polycycloheteroalkylenealkylene, aryleneheteroalkylene, heteroarylenealkylene, aryleneheteroalkenylene, heteroarylenealkenylene groups. The heterohydrocarbylene groups may be attached to the remainder of the compound by any carbon or hetero (*e.g.* nitrogen) atom.

Where reference is made to a carbon atom of a hydrocarbyl or other group being replaced by an O, S, Se or N atom, what is intended is that:



–CH= is replaced by –N=; or

$-CH_2-$ is replaced by –O–, –S– or –Se–.

The term 'heteroalkyl' includes alkyl groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N.

The term 'heteroalkenyl' includes alkenyl groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N.

The term 'heteroalkynyl' includes alkynyl groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N.

The term 'heteroaryl' includes aryl groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N. Preferred heteroaryl are C_{5-14}heteroaryl. Examples of heteroaryl are pyridyl, pyrrolyl, thienyl or furyl.

Other examples of heteroaryl groups are monovalent derivatives of acridine, carbazole, β-carboline, chromene, cinnoline, furan, imidazole, indazole, indole, indolizine, isobenzofuran, isochromene, isoindole, isoquinoline, isothiazole, isoxazole, naphthyridine, perimidine, phenanthridine, phenanthroline, phenazine, phthalazine, purine, pyran, pyrazine, pyrazole, pyridazine, pyridine, pyrimidine, pyrrole, pyrrolizine, quinazoline, quinoline, quinolizine, quinoxaline, thiophene and xanthene. Preferred heteroaryl groups are five- and six-membered monovalent derivatives, such as the monovalent derivatives of furan, imidazole, isothiazole, isoxazole, pyran, pyrazine, pyrazole, pyridazine, pyridine, pyrimidine, pyrrole, pyrrolizine and thiophene. The five-membered monovalent derivatives are particularly preferred, *i.e.* the monovalent derivatives of furan, imidazole, isothiazole, isoxazole, pyrazole, pyrrole and thiophene.

The term 'heteroalkylene' includes alkylene groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N.

The term 'heteroalkenylene' includes alkenylene groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N.

The term 'heteroalkynylene' include alkynylene groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N.

The term 'heteroarylene' includes arylene groups in which up to three carbon atoms, preferably up to two carbon atoms, more preferably one carbon atom, are each replaced independently by O, S, Se or N, preferably O, S or N. Preferred heteroarylene are C_{5-14}heteroarylene. Examples of heteroarylene are pyridylene, pyrrolylene, thienylene or furylene.

Other examples of heteroarylene groups are divalent derivatives (where the valency is adapted to accommodate the q instances of the linker L_M) of acridine, carbazole, β-carboline, chromene, cinnoline, furan, imidazole, indazole, indole, indolizine, isobenzofuran, isochromene, isoindole, isoquinoline, isothiazole, isoxazole, naphthyridine, perimidine, phenanthridine, phenanthroline, phenazine, phthalazine, purine, pyran, pyrazine, pyrazole, pyridazine, pyridine, pyrimidine, pyrrole, pyrrolizine, quinazoline, quinoline, quinolizine, quinoxaline, thiophene and xanthene. Preferred heteroarylene groups are five- and six-membered divalent derivatives, such as the divalent derivatives of furan, imidazole, isothiazole, isoxazole, pyran, pyrazine, pyrazole, pyridazine, pyridine, pyrimidine, pyrrole, pyrrolizine and thiophene. The five-membered divalent derivatives are particularly preferred, *i.e.* the divalent derivatives of furan, imidazole, isothiazole, isoxazole, pyrazole, pyrrole and thiophene.

Substitution

A is independently a substituent, preferably a substituent S_{ub}^1. Alternatively, A may be ^2H.

S_{ub}^1 is independently halogen, trihalomethyl, $-NO_2$, $-CN$, $-N^+(R^1)_2O^-$, $-CO_2H$, $-CO_2R^1$, $-SO_3H$, $-SOR^1$, $-SO_2R^1$, $-SO_3R^1$, $-OC(=O)OR^1$, $-C(=O)H$, $-C(=O)R^1$, $-OC(=O)R^1$, $-NR^1_2$, $-C(=O)NH_2$, $-C(=O)NR^1_2$, $-N(R^1)C(=O)OR^1$, $-N(R^1)C(=O)NR^1_2$, $-OC(=O)NR^1_2$, $-N(R^1)C(=O)R^1$, $-C(=S)NR^1_2$, $-NR^1C(=S)R^1$, $-SO_2NR^1_2$, $-NR^1SO_2R^1$, $-N(R^1)C(=S)NR^1_2$, $-N(R^1)SO_2NR^1_2$, $-R^1$ or $-Z^1R^1$.

Z^1 is O, S, Se or NR^1.

R^1 is independently H, C_{1-8}hydrocarbyl, C_{1-8}hydrocarbyl substituted with one or more S_{ub}^2, C_{1-8}heterohydrocarbyl or C_{1-8}heterohydrocarbyl substituted with one or more S_{ub}^2.

S_{ub}^2 is independently halogen, trihalomethyl, $-NO_2$, $-CN$, $-N^+(C_{1-6}alkyl)_2O^-$, $-CO_2H$, $-CO_2C_{1-6}alkyl$, $-SO_3H$, $-SOC_{1-6}alkyl$, $-SO_2C_{1-6}alkyl$, $-SO_3C_{1-6}alkyl$, $-OC(=O)OC_{1-6}alkyl$, $-C(=O)H$, $-C(=O)C_{1-6}alkyl$, $-OC(=O)C_{1-6}alkyl$, $-N(C_{1-6}alkyl)_2$, $-C(=O)NH_2$, $-C(=O)N(C_{1-6}alkyl)_2$, $-N(C_{1-6}alkyl)C(=O)O(C_{1-6}alkyl)$, $-N(C_{1-6}alkyl)C(=O)N(C_{1-6}alkyl)_2$, $-OC(=O)N(C_{1-6}alkyl)_2$, $-N(C_{1-6}alkyl)C(=O)C_{1-6}alkyl$, $-C(=S)N(C_{1-6}alkyl)_2$, $-N(C_{1-6}alkyl)C(=S)C_{1-6}alkyl$, $-SO_2N(C_{1-6}alkyl)_2$, $-N(C_{1-6}alkyl)SO_2C_{1-6}alkyl$, $-N(C_{1-6}alkyl)C(=S)N(C_{1-6}alkyl)_2$, $-N(C_{1-6}alkyl)SO_2N(C_{1-6}alkyl)_2$, $C_{1-6}alkyl$ or $-Z^1C_{1-6}alkyl$.

Where reference is made to a substituted group, the substituents are preferably from 1 to 5 in number, most preferably 1.

However, molecular mass labels of the invention will generally comprise 1 or more, typically between 1 and 100 (*e.g.* 1 to 50, preferably 1 to 20) substituents S_{ub}^1 or S_{ub}^2, typically F or I, in order to vary the masses of the molecular mass labels.

Preferred examples of substituent A are shown in figure 14.

Miscellaneous

A may optionally be a monovalent dendrimer radical or a monovalent dendrimer radical substituted with one or more substituents S_{ub}^1.

General

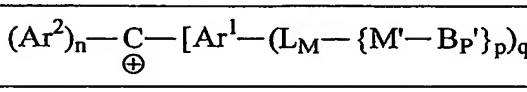
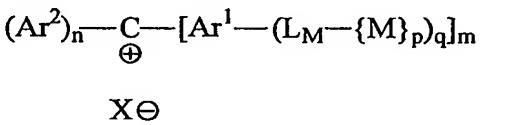
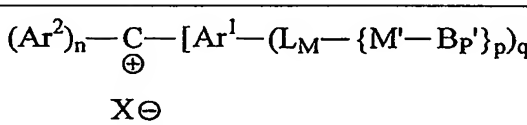
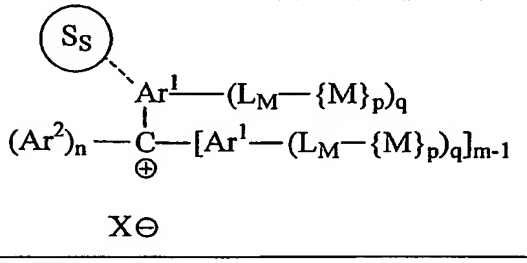
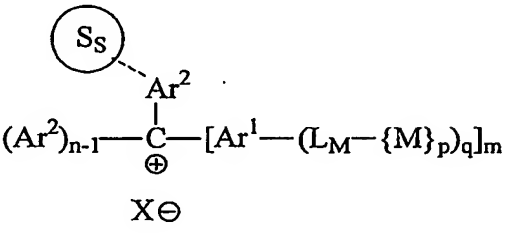
The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The term "about" in relation to a numerical value *x* means, for example, $x \pm 10\%$.

The word "substantially" does not exclude "completely" *e.g.* a composition which is "substantially free" from Y may be completely free from Y. Where necessary, the word "substantially" may be omitted from the definition of the invention.

Tables

Table 1 — C★ is a cation

Formula	Structure
Formula (I)	
Formula (IIb)	
Formula (IIIb)	
Formula (IVbii)	
Formula (IVbiii)	

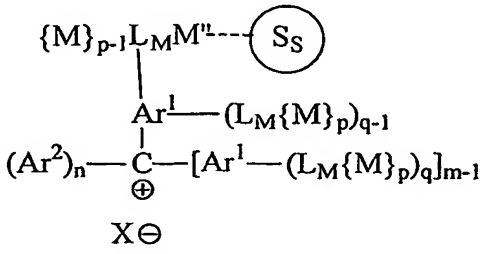
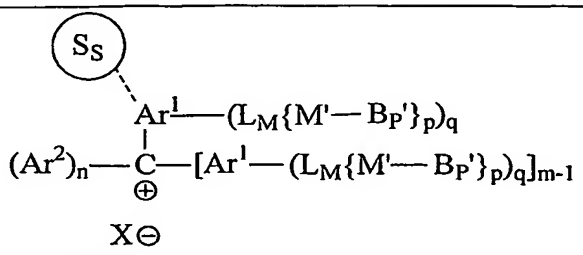
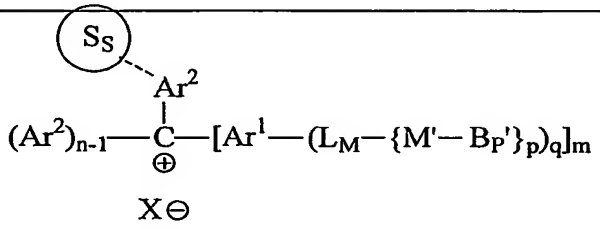
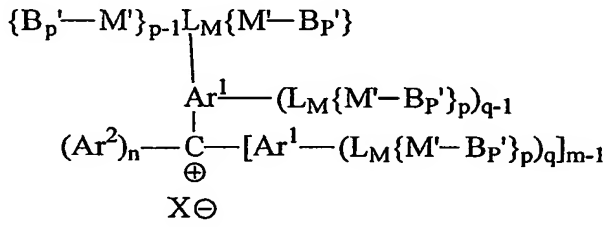
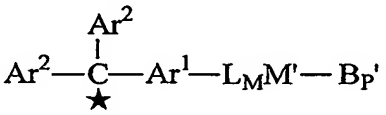
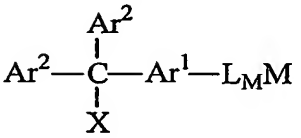
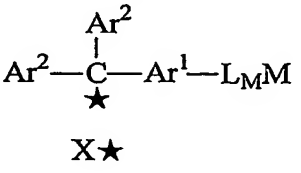
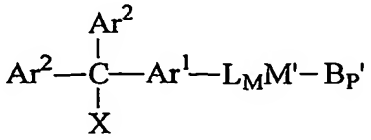
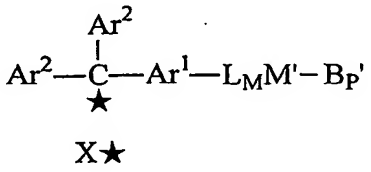
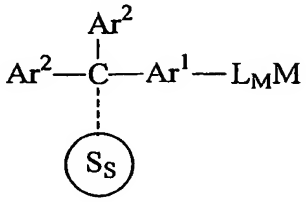
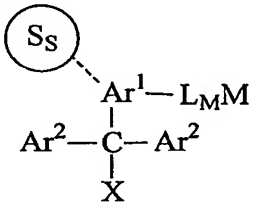
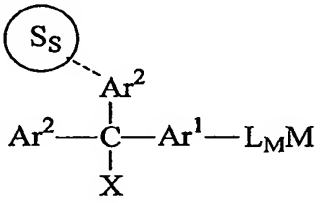
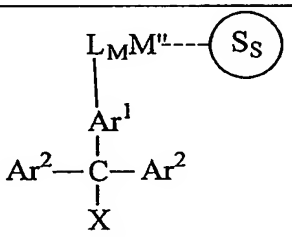
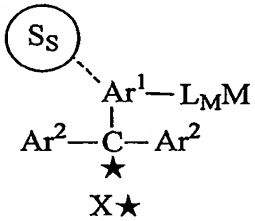
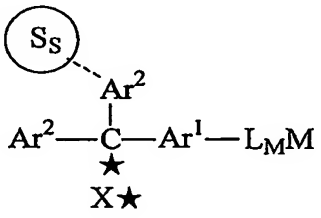
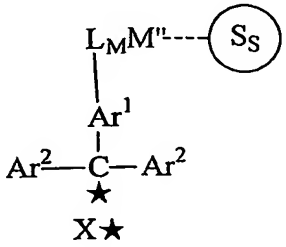
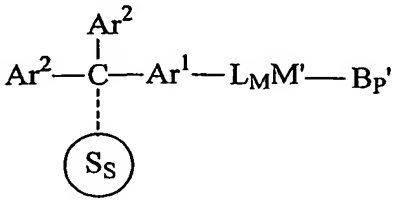
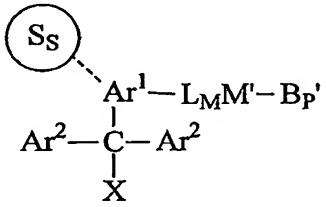
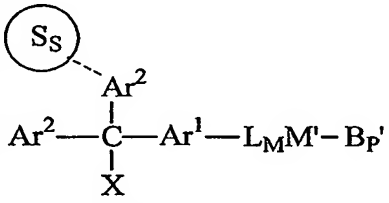
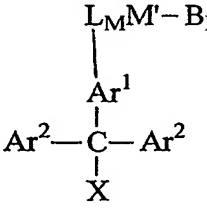
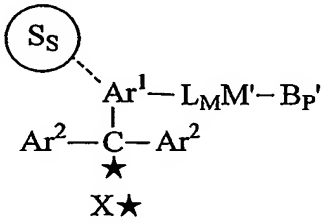
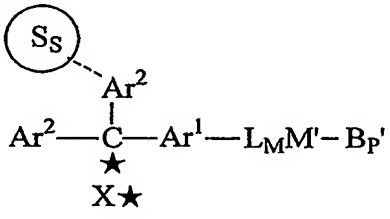
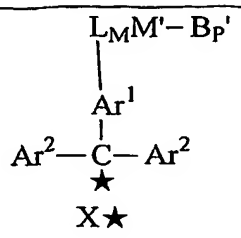
Formula (IVbiv)	
Formula (Vbii)	
Formula (Vbiii)	
Formula (Vbiv)	

Table 2 — n = 2, m = 1, p = 1 and q = 1

Formula	Structure
Formula (I)	
Formula (IIa)	
Formula (IIb)	

Formula (IIIa)	
Formula (IIIb)	
Formula (IVai)	
Formula (IVaii)	
Formula (IVaiii)	
Formula (IVaiv)	
Formula (IVbii)	
Formula (IVbiii)	

Formula (IVbiv)	
Formula (Vai)	
Formula (Vaii)	
Formula (Vaiii)	
Formula (Vaiv)	
Formula (Vbii)	
Formula (Vbiii)	

Formula (Vbiv)	

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 demonstrates conceptually the effect of the signal on a mass spectrum by a compound of formula (IIa) or (IIb) of the invention. Free biopolymer, such as a peptide, has poorer desorption properties characterised by a smaller peak on the left of mass-spectrum whereas desorption improves when the same molecule is conjugated to a compound of the invention.

Figure 2 shows the steps of biopolymer with a compound of formula (IVai). The derivativisation of a biopolymer with a compound of the invention can be carried out more conveniently by utilising the solid phase-based format, whereby the compound is temporarily covalently attached to a solid support. This eliminates all the separation steps associated with homogenous approach as the only additional step required would be a washing step. The solid support can be a resin, a surface or a porous surface. Alternatively, the solid support may be a mass-spectrometry sample plate, which dramatically decreases the sample preparation time. Both gold, glass- and plastic-coated plates are compatible with this approach.

Figure 3 shows the steps of 'reverse' biopolymer derivativisation on a covalent solid support whereby the release of the biopolymer derivative happens simultaneously with the derivativisation process. The process is applicable M groups involving leaving groups.

Figure 4 shows the steps of biopolymer derivativisation on an ionic solid support.

Figure 5 shows of the steps of solid support-assisted biopolymer derivativisation. The biopolymer is first trapped onto a solid support and then labelled with a compound of formula (IIa) or (IIb). An advantage of this technique is that a preliminary sample enrichment occurs, since not all of the biopolymer in the sample will stick to the solid support surface.

Figure 6 shows the mass spectrum obtained when analysing an Gly-Gly-O-acyl dipeptide conjugated with a trityl compound of the invention.

Figure 7 shows the mass spectrum obtained when analysing a conjugate of a peptide with a trityl compound of the invention.

Figure 8 compared the mass spectra of a BSA digest without (8A) and with (8B) labelling.

Figure 9 shows the mass spectrum obtained when analysing a mixture of trityl-labelled amines.

Figures 10A and 10B show preferred examples of group L_M.

Figures 11A and 11B show preferred examples of group M.

Figures 12A and 12B shows preferred examples of groups Ar^1 and Ar^2.

Figure 13 shows preferred examples of groups X and X★.

Figure 14 shows preferred examples of substituent group A.

MODES FOR CARRYING OUT THE INVENTION

Materials and Methods

The solid supports were Tenta Gel Macrobeads OH and NH_2, 280-320 microns, Rapp Polymer. (MA)LDI-TOF mass-spectra were recorded on a PE-ABI Voyager™ Elite Reflectron Delayed Extraction Instrument. TLC were carried out with Merck silica gel (Kieselgel 60 F_{254} precoated plates and Kieselgel 60 0.040–0.063 mm). HPLC was carried out on a Waters system (Milford, MA, USA). Phosphoroamidite couplings were carried out in an ABI 394 DNA/RNA synthesiser. Chemicals and solvents were from Sigma/Aldrich/Fluka (USA), and BDH/Merck.

Example 1 — Conjugation of a trityl tag (in solution phase) with solid support-bound biopolymer

A 15mer poly-T oligonucleotide was synthesised on an ABI 394 DNA synthesiser using a T CPG support according to standard protocols of phosphoramidite chemistry on 0.2 μmol scale. After the last coupling, a MMTr-protected 'aminolink' phosphoramidite (Glen Res., USA) was added to a growing chain and deprotected using standard deblocker (2% DCA in DCM). The column was removed from the synthesiser, and after 10 min wash with acetonitrile it was attached to two 5 ml syringes and washed with a 0.1M solution of NHS-activated 4,4'-dimethoxy-4"-carboxyethyl trityl for 10 min at RT. The column was then washed with (3 x 10 ml) acetonitrile, placed on a DNA synthesiser and deprotected with ammonia according to standard protocols. The residue obtained after the evaporation was dissolved in 0.1 ml of 2M $LiClO_4$ and precipitated from cold acetone (1.5 ml). The precipitate was washed with 0.5ml of acetone and dried.

Example 2 — Homogenous conjugation of a trityl with non-polymeric ligands

A solution of NHS-activated 4,4'-dimethoxy-4"-carboxyethyl trityl (0.1M) in THF/dioxane (1:1) was mixed with a solution (0.5-1M) of an amine or of a mixture of amines (for example, propyl amine, butyl amine, pentyl amine, hexyl amine and phenethyl amine), typically 10 ml of a solution of an activated trityl with 5 ml of an amine solution. The mixtures were purified on prep-TLC (2mm-thick glass plates with UV254 indicator, Analtech/Aldrich-Sigma), typically in chloroform with 0.5% triethylamine. The areas containing the desired products were scratched off the plate, and the conjugates or the mixtures thereof were eluted using same solvent with 2-5% MeOH, filtered through a layer of glass wool, evaporated and dried.

Example 3 — Homogenous conjugation of a nhs-activated trityl with polymeric ligands

A peptide, an oligonucleotide, or any other biopolymer containing a (primary) amino group, is dissolved in a mixture of water and acetonitrile depending on its solubility, typically 20-50% of water in CH_3CN. Non-aminogroup-containing buffers (ie. 50 mM sodium phosphate, 0.15 M NaCl, pH 7.2, or a bicarbonate buffer, but an additional desalting step may then need to be introduced to cut

off the metal ions prior to mass-spectrometry) can be used to keep the pH at between 7-9. For particularly poorly soluble ligands other solvents may be used such as THF, DMSO, etc.

A solution of NHS-activated 4,4'-dimethoxy-4"-carboxyethyl trityl in acetonitrile or THF is added in approx. 5-10 times excess compared to an amine component. Conjugation usually reaches the maximum yield over 2-4 hours of reaction time. The conjugate formed can be analysed by MS directly, or after HPLC-purification.

Example 4 — Conjugation of a solid phase-immobilised nhs-activated trityl tag with a ligand

A Solid Phase-Immobilised NHS-Activated Trityl Tag was prepared by either method 1 or method 2.

Method 1: A NHS-Activated 4,4'-dimethoxy-4"-carboxyethyl trityl tag was covalently attached to hydroxyl groups of 200 μm Rapp Polymer beads by shaking the suspension of 100 mg of the resin in 5 ml of 0.1 M solution of trityl chloride tag in dry pyridine at +4°C for 3 hours and then washing the resin with pyridine and acetonitrile and drying *in vacuo*.

Method 2. A 5'-tritylated thymidine phosphoramidite was prepared from NHS-activated 4,4'-dimethoxy-4"-carboxyethyl trityl chloride in a standard way [M.J. Gait, Oligonucleotide Synthesis: A Practical Approach, IRL, Oxford, 1984]. The Rapp Polymer beads (2 x 40 mg) were placed in two 1 micromol scale DNA synthesis columns (Glen Res.). The first column was coupled with the said phosphoramidite on an ABI DNA synthesiser using manual supply of reagents (0.1M solution of a phosphoramidite and other standard phosphoramidite synthesis reagents) with a coupling step of 15 min. The second column was first derivatised with a trebler phosphoramidite (Glen Res.) according to the manufacturer's protocols and then coupled with the trityl tag-containing phosphoramidite as described for the first column. Both columns were excessively washed with acetonitrile.

The trityl loading of the solid supports produced by either method was determined spectrophotometrically (absorbance measurements at 490nm) to be 0.21 mmol/g for a straight attachment and 0.39 mmol/g for a tritylation on top of the trebling synthon. (The hydroxyl group loading of the Rapp polymer used was 0.25mmol/g).

To the solid support prepared as described above, a mixture of compounds to be labelled (typically peptides) is added, typically in a mixture of 20-50% water in acetonitrile. After incubation, with occasional shaking, for 60-120 min the resin is washed with several volumes of the same solvent, and the conjugated products are cleaved off the resin, typically by adding 0.5-2% TFA in appropriate solvent. The collected sample is then analysed by MS.

Example 5 — Mass spectrometry analysis of a derivatised Gly-Gly dipeptide

Figure 6 shows the mass spectrum obtained from a compound of the invention comprising a derivatised Gly-Gly-O-acyl dipeptide biopolymer.

The ion of formula (I) containing the derivatised Gly-Gly-O-acyl biopolymer is observed at the peak at molecular weight 516.5. There was no peak corresponding to the free dipeptide.

The fragment of formula (VI), in which the derivatised Gly-Gly-O-acyl biopolymer has been lost, is observed at the peak at the molecular weight 374.6.

Example 6 — Mass spectrometry analysis of a derivatised peptide

Figure 7 shows the mass spectrum obtained from a compound of the invention comprising a derivatised peptide biopolymer. The free peptide had a molecular weight of 310.

The ion of formula (I) containing the derivatised peptide biopolymer is observed at the peak at molecular weight 665.0.

The fragment of formula (VI), in which the derivatised peptide has been lost, is observed at the peak at the molecular weight 375.0.

Significantly, there is only a very small peak at molecular weight 310, where a peak corresponding to the free biopolymer would be found. The relative size of the peaks at 665.0 and 310 thus demonstrate the significantly improved ionisability of the compounds of the invention compared with free biopolymer.

Example 7 — Spectral improvement by trityls

Three proteins (BSA, β-casein and ADH) were digested with trypsin and the resulting peptides analysed by MALDI-TOF mass spectrometry with or without derivatisation. The number of peptides identified for each protein is shown below. The theoretical total number of peptides that would be produced by trypsin digestion of each protein was calculated *in silico* and is shown in the second column the table below.

Protein	Number of theoretical peptides+	Total number of peptides identified		MASCOT search score*	
		Underivatised	Derivatised	Underivatised	Derivatised
BSA	144	14 (10%)	41 (28%)	132	126
β-casein	27	4 (15%)	13 (48%)	no match	123
ADH	60	7 (12%)	18 (30%)	77	111

\+ The number of theoretical peptides for each protein was generated assuming one missed cleavage and disregarding di- and mono-amino acids generated.

* Score is -10*Log(P), where P is the probability that the observed match is a random event. Protein scores greater than 63 are significant ($p<0.05$).

Derivatisation of peptides with trityl groups of the invention thus improves detection, as a significantly larger number of peptides was detected for each of the three proteins when derivatisation was used. Furthermore, protein identification by mass fingerprinting can be improved.

Taking β-casein as an example, the number of detectable fragments more than tripled, and the derivatised spectrum allowed a MASCOT-based identification which was not previously possible.

Example 8 — BSA fragmentation and mass spectrometry

Bovine serum albumin (BSA) was digested with trypsin and analysed by MALDI-TOF. The resulting spectrum is shown in Figure 8A. The experiment was repeated, but the peptide mixture was labelled

with a dimethoxytrityl label after trypsin digestion. The spectrum in Figure 8B shows the dramatic increase in visible ions due to the trityl label. Four specific peptides have been highlighted in both spectra.

Example 9 — Mass spectrometry of amines

A solution of NHS-activated 4,4'-dimethoxy-4"-carboxyethyl trityl (0.1M) in THF/dioxane (1:1) was mixed with a solution (0.5-1M) of an amine or of a mixture of amines (for example, propyl amine, butyl amine, pentyl amine, hexyl amine and phenethyl amine), typically 10 ml of a solution of an activated trityl with 5 ml of an amine solution. The mixtures were purified on prep-TLC (2mm-thick glass plates with UV254 indicator, Analtech/Aldrich-Sigma), typically in chloroform with 0.5% triethylamine. The areas containing the desired products were scratched off the plate, and the conjugates or the mixtures thereof were eluted using same solvent with 2-5% MeOH, filtered through a layer of glass wool, evaporated and dried. Figure 9 shows a spectrum obtained in this way.

It will be understood that the invention is described above by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.